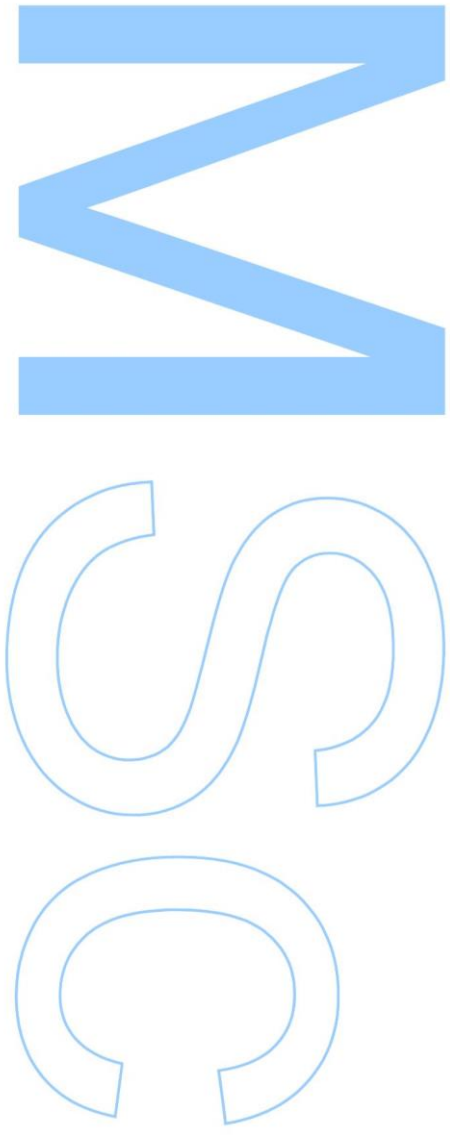
Unravelling molecular mechanisms of *polo* alternative polyadenylation

Mafalda Sofia Bastos Pinho
Mestrado em Biologia Celular e Molecular
Departamento de Biologia
Faculdade de Ciências da Universidade do Porto
2014

Orientador
Alexandra Moreira, PhD, IBMC-UP

Coorientador
Jaime Freitas, PhD, IBMC-UP



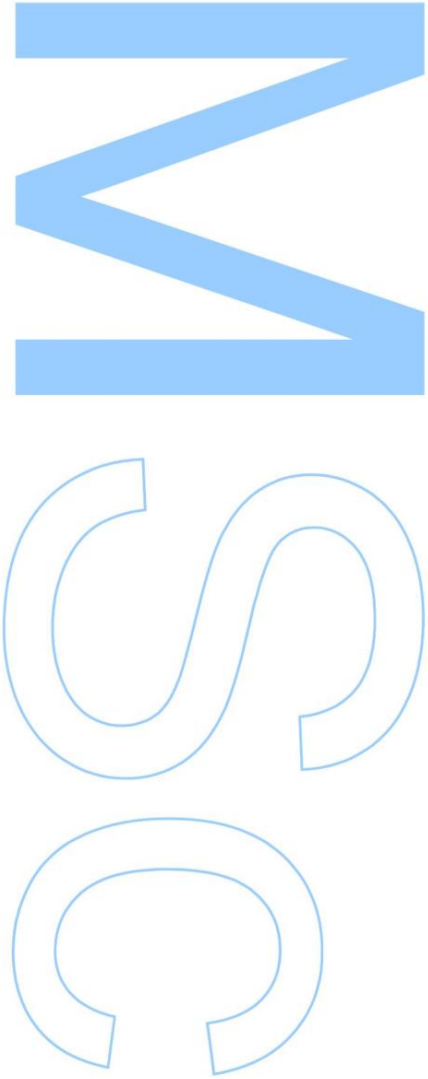
IBMC
INSTITUTO DE BIOLOGIA MOLECULAR E CELULAR
INSTITUTE FOR MOLECULAR AND CELL BIOLOGY

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Dissertação para a candidatura ao grau de Mestre em Biologia Celular e Molecular submetida à Faculdade de Ciências da Universidade do Porto.

A presente dissertação foi orientada pela Doutora Alexandra Moreira e foi realizada no Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto.

"We are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress."

Richard P. Feynman

I have a friend who's an artist and has sometimes taken a view which I don't agree with *very well. He'll hold up a flower and say "look how beautiful it is," and I'll agree. Then he says "I as an artist can see how beautiful this is but you as a scientist take this all apart and it becomes a dull thing," and I think that he's kind of nutty. First of all, the beauty that he sees is available to other people and to me too, I believe, although I might not be quite as refined aesthetically as he is, I can appreciate the beauty of a flower.*

At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes. The fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement, the mystery and the awe of a flower. It only adds. I don't understand how it subtracts.

Richard P. Feynman

Acknowledgements/Agradecimentos

Em primeiro lugar, gostaria de agradecer à Doutora Alexandra Moreira, minha orientadora, por me ter aceitado como sua aluna de mestrado. E por me ter conhecido nos dias de entusiasmo, em que descobrir o mundo ficava ao alcance de uma experiência, assim como nos dias menos bons. Obrigada por todos os incentivos que me fizeram avançar e Querer estar na ciência pelo simples prazer de me entusiasmar e querer fazer do laboratório uma casa. Muito obrigada pelo entusiasmo e dedicação que sempre demonstrou, e que estiveram na génese de todo o meu empenho. Um último obrigado por ter contribuído para o meu crescimento, não só como cientista mas também como pessoa. Um abraço apertado.

Um grande obrigado a todo o *Gene Regulation*, por todos os momentos partilhados no laboratório, e também nas *happy hours*!!! A ciência sem vocês seria muito mais triste.

À Ana, por me fazer sentir uma exploradora, quando saíamos à descoberta dos cantos mais recônditos do instituto. Obrigada pela tua amizade.

À Inês Boal, por todas as conversas científicas e, mais ainda, por todas as outras.

Um obrigado especial à minha Baldi, que não está cá… Eu nunca me esqueço de ti. Obrigada por toda a ajuda que sempre me deste e por me fazeres entender que o mais importante está mesmo aqui ao nosso lado. Obrigada pela AMIZADE que temos. Mesmo estando longe, conseguimos fazer parte da vida uma da outra como se Brighton ficasse na Rua do Campo Alegre.

Ao padrinho, sempre presente.

À Mariana Osswald, pela disponibilidade e ajuda técnica.

À Augusta, ao Torcato e ao Pedro Pinto por me introduzirem este maravilhoso mundo e por me esclarecerem todas as dúvidas.

Obrigada à Doutora Carla Lopes por estar sempre disposta a ajudar e por me mostrar como “os sorrisos fazem o trabalho parecer muito mais fácil”.

Obrigada à Cristina por todo o carinho e pelas palavras adequadas. Gosto muito de si e desejo que continue a ser Pessoa.

Obrigada à Carla, estou desde os 3 anos a agradecer-te. Já não há mais palavras que nos sirvam.

Obrigada ao Jaime por ter estado sempre presente para me ajudar a resolver todos os contratempos, grandes ou pequenos, a qualquer hora do dia. Obrigada por me teres dedicado fins-de-semana e tudo! És um ótimo mentor, tanto a nível profissional como pessoal. Deixo-te um enorme, enorme, enorme obrigado. “*Give a man a Truth, and he will*

think for a day. Teach a man to reason, and he will think for a lifetime!" Este é um dos ensinamentos que levarei comigo para onde quer que vá. Obrigada.

À minha família, por me ter aturado ainda mais mal disposta e stressada do que o costume, nos últimos tempos. Prometo que vos pago um jantar!

Ao João, porque todas as palavras do mundo não seriam suficientes para te agradecer aquilo que representas para mim. Obrigada por seres incansável e por saberes sempre dar-me todo o apoio nos momentos mais acertados. Por estares sempre "ao" e "do" meu lado, incondicionalmente.

Resumo

A poliadenilação alternativa (APA) é um fenómeno amplamente conservado que contribui para a complexidade do transcriptoma ao dar origem a múltiplas isoformas de RNA mensageiro (mRNA) que diferem na sua sequência codificante ou na sua região 3' não traduzida (UTR). O gene regulador do ciclo celular, *polo*, usa dois sinais de poliadenilação (pA1 e pA2) presentes no 3'UTR para produzir dois mRNAs distintos, que possuem diferentes eficiências de tradução. Foi anteriormente mostrado, em *Drosophila*, que organismos sem o sinal pA2 não produzem níveis suficientes de proteína Polo, o que provoca falhas na proliferação dos histoblastos abdominais durante a metamorfose, comprometendo o desenvolvimento do organismo.

Com vista a perceber os mecanismos por trás do uso do sinal distal, procuraram-se, na região 3' não traduzida, elementos reguladores putativos com importância na formação da extremidade 3' do mRNA. Foram encontradas três sequências conservadas cuja composição nucleotídica e posição se assemelhavam a upstream sequence elements (USEs). Uma dessas sequências despertou particular interesse devido ao seu elevado conteúdo em timina. Esta sequência mostrou ser conservada, quando comparadas duas espécies de *Drosophila* distantes do ponto de vista evolutivo, e semelhante às sequências dos USEs de outros genes previamente descritos. Através de ensaios de clivagem e poliadenilação *in vitro*, o USE1 mostrou ser importante na poliadenilação do sinal proximal, pA1. Contudo, ainda não havia sido demonstrada uma função para esta sequência *in vivo*. O presente trabalho demonstra, *in vivo*, que o USE1 tem um papel na formação da extremidade 3' usando o sinal proximal pA1 do gene *polo*.

Foi previamente mostrado que as proteínas HuR e PTB, duas proteínas de ligação ao RNA envolvidas em mecanismos de processamento do pre-mRNA, se ligam à sequência USE1 do pre-mRNA do *polo.* Neste trabalho, a principal questão a responder é se as proteínas HuR/Elav e Heph/PTB têm uma função na escolha do sinal de poliadenilação usado. Usando dsRNAs para o silenciamento dos genes *elav* e *heph* em células Schneider's 2 de *Drosophila*, observou-se que o silenciamento destes genes leva a um aumento do uso do sinal pA1 e a uma diminuição no uso do sinal pA2. Estes resultados identificam a Elav e Heph como moduladores da poliadenilação alternativa do *polo*, uma vez que modificações nos níveis destas proteínas levam à alteração do padrão de poliadenilação do gene *polo.*

Palavras-chave: 3'UTR; USE; formação da extremidade 3'; proteínas de ligação ao RNA; silenciamento mediado por RNAi; poliadenilação alternativa; *polo*; *Drosophila melanogaster*

Abstract

Alternative polyadenylation (APA) is a widespread conserved phenomenon that contributes to the complexity of the transcriptome by giving rise to multiple messenger RNA (mRNA) isoforms that differ in their coding sequence or in their 3' untranslated region (UTR). The cell cycle regulator *polo* uses two polyadenylation (pA) signals in the 3'UTR (pA1 and pA2) to produce two distinct mRNAs with different translation efficiencies. It had been previously shown that flies without pA2 do not produce sufficient Polo protein and abdominal histoblasts fail to proliferate during metamorphosis, compromising the organism development.

In order to understand the mechanisms behind the usage of the distal pA site, 3'UTR was searched for putative regulatory elements important for mRNA 3' end formation. Three conserved sequences were found whose nucleotide composition and position resembled USEs. One of these sequences seemed particularly interesting because of its high thymine deoxynucleotide content. This sequence is conserved between distantly related *Drosophila* species and similar to several described USEs. Therefore, the *in vitro* role of USE1 was addressed but its *in vivo* function had yet to be identified. Here we show that USE1 has an *in vivo* role in *polo* mRNA 3' end formation at the proximal pA1 site.

polo USE1 was shown to bind HuR and PTB, RNA-binding proteins that are involved in several pre-mRNA processing mechanisms. We have now asked if HuR/Elav and PTB/Heph have a function in *polo* pA site selection. Using double-stranded RNA (dsRNAs) for RNAi knockdown of *elav* and *heph* causes an increase in pA1 site usage and a decrease in pA2 levels. Taken together our results identify Elav and Heph as modulators of *polo* APA, since changes in their levels lead to alterations in the *polo* polyadenylation pattern.

Keywords: 3'UTR; USE; 3' end formation; RNA-binding proteins; RNAi-mediated silencing; alternative polyadenylation; *polo; Drosophila melanogaster*

Table of Contents

Tables list

Figures list

Abbreviations

µg – microgram;

µl – microliter;

7 meG – 7-methylguanosine;

A – adenine;

A_{260} – absorbance at 260 nanometers;

A_{280} – absorbance at 280 nanometers;

APA – alternative polyadenylation;

C – cytosine;

C/P – cleavage/polyadenylation;

CaMV – cauliflower mosaic virus;

CAT – chroramphenicol acetyltranferase;

cDNA - complementary deoxyribonucleic acid;

CFI – cleavage factor I_m;

COX – cyclooxygenase;

CPSF – cleavage and polyadenylation specificity factor;

CstF – Cleavage and stimulation Factor;

C_T – threshold cycle;

CTD - carboxy terminal domain;

ddH_2O - distilled deionized water;

DNA – deoxyribonucleic acid;

DNase – deoxyribonuclease;

dNTPs – deoxyribonucleotide;

DSE – downstream sequence element;

DsRed – red fluorescent protein;

dsRNA – double-stranded ribonucleic acid;

DTT – dithiothreitol;

E – eficiency;

EDTA – ethylenediaminetetraacetic acid;

elav – embryonic lethal abnormal vision;

EST – expressed sequence tag;

FBS – fetal bovine serum;

Fne – found in neurons;

g – grams;

G – guanine;

gfp – green fluorescent protein;

heph - *Hephaestus*;

HIV-1 – human immunodeficiency virus type 1;

hnRNP – heterogeneous nuclear ribonucleoproteins;

HuR – ; Hu-antigen R;

Ig – immunoglobulin;

LTR – long terminal repeats;

min – minutes;

ml – milliliters;

MLTU – adenovirus major late transcription unit;

mM – millimolar;

mRNA – messenger RNA;

nm – nanometers;

ºC – celsius degrees;

pA – polyadenylation;

PABPN 1 – poly (A) binding protein nuclear 1;

PAP – poly (A) Polymerase;

PCR - polymerase chain reaction;

PBD– polo-box domain;

pre-mRNA – precursor messenger ribonucleic acid;

PTB – polypyrimidine tract-binding protein;

qPCR – quantitative real-time polymerase chain reaction;

RACE – rapid amplification of cDNA ends;

RBP – ribonucleic acid-binding proteins;

rcf – relative centrifugal force;

RNA – ribonucleic acid;

RNA pol II – ribonucleic acid polymerase II;

RNAi – ribonucleic acid interference;

RNase – ribonuclease;

RT-PCR – reverse transcription polymerase chain reaction;

S2 - Schneider's *Drosophila* line 2;

sec – seconds;

SV40 – Simian virus 40;

U – uracil;

USE – upstream sequence element;

UTR – untranslated region;

Wt – wild-type.

I. Introduction

1. Key players in mRNA 3'end formation

The main processes involved in maturation of nascent mRNAs in high eukaryotes are capping, splicing and polyadenylation. Capping involves the addition of a 7-methylguanosine (7 meG) cap at the 5' end of the precursor mRNA (pre-mRNA) and splicing consists in the exclusion of introns and the assembly of the exons (reviewed in (1)). Regarding mRNA 3' end formation, it refers to a two-step process that includes 3' end cleavage and subsequent polymerization of a string of adenosines to the pre-mRNA (2). The addition of a polyadenylation (pA) tail is known to be a co-transcriptional process (3) and it plays an important role in several processes such as mRNA transport from the nucleus to the cytoplasm and the subcellular localization of specific mRNAs, its stability and translation in certain developmental stages (reviewed in (2)). The polyadenylation (pA) site is defined by multiple *cis*-elements and its recognition relies on several multi-subunit protein complexes, the *trans*-acting factors, involved in 3' end formation. In the following sections, both *cis*-acting elements and *trans*-acting factors will be discussed.

A. *cis*-acting elements required for mRNA 3'end formation

Specific sequences on the precursor RNA play a crucial role in determining the efficiency of mRNA 3' end formation in a given cellular environment. The *cis*-acting elements involved in this process consists of an AAUAAA hexamer or a variant of this, a U- or GU-rich downstream element (DSE), and other auxiliary sequences including the U-rich upstream sequence element (USE) (Figure 1).

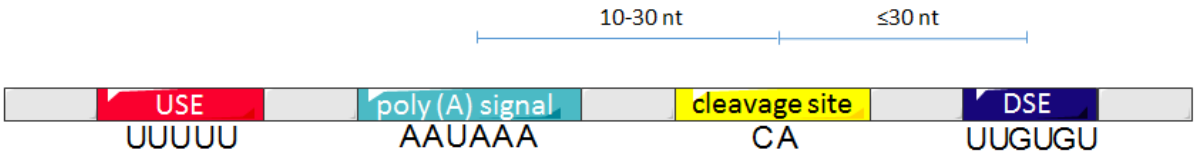


Figure 1. Representation of *cis*-acting elements involved in mRNA 3'end formation. U-rich and U/GU-rich sequences that are typically found upstream and downstream of the pA signal are named USE (upstream sequence element) and DSE (downstream sequence element), respectively. The sequence AAUAAA is the most conserved pA signal in both humans and *Drosophila* and the cleavage site is typically defined by a CA dinucleotide. Adapted from (2).

The first signal to be recognized by the cleavage/polyadenylation (C/P) machinery is the pA signal, AAUAAA or a variant of this hexamer. The pA signal AAUAAA was first identified in six different mRNAs by Proudfoot and Brownlee (4) and was shown to be present in the majority of the genes and to have a relevant role in mRNA processing both *in*

vivo and *in vitro* (5-7). While AAUAAA is the most conserved sequence, AUUAAA is the most common variation of this canonical signal both in humans and *Drosophila* (8-10). Although there are several other variants in the genome, they are present in a very low frequency and demonstrate a strong reduction on the efficiency of cleavage and polyadenylation (6, 11). The hexamer is necessary for both the cleavage and the polyadenylation reactions (6, 7, 11) because it binds important protein factors of the cleavage and polyadenylation (C/P) machinery.

Downstream sequence elements are also important to determine the efficiency of a pA site because of their importance to the cleavage reaction. The best characterized auxiliary downstream element is a G-rich sequence of the Simian virus 40 (SV40) late mRNA (12). Deletions as well as point mutations in this G-rich sequence lead to a decrease in the polyadenylation efficiency (13). Despite being present in the majority of the genes (14), the motif itself is poorly conserved, and its main function has been assigned to stabilize the binding of a multiple sub-unit complex of the C/P machinery, the cleavage and stimulation factor (CstF) (15).

Before the addition of a pA tail the pre-mRNA is cleaved. The cleavage site and its efficiency are influenced by the position of the pA signal relative to a DSE (16). The site of cleavage is determined by a single nucleotide preference in the following order: A > U > C > G (16). Typically, endonucleotic cleavage occurs at a CA dinucleotide and studies with thrombophilia have shown that pro-thrombin mRNA 3' end formation efficiency is related to this dinucleotide. Changes in the dinucleotide affect the efficiency of the reaction and lead to alterations in the amount of mRNA and protein produced (17, 18). Briefly, a change in the CG dinucleotide to either TG (18) or CA (17), increases dramatically the levels of pro-thrombin, increasing the risk of developing thrombosis.

The best characterized auxiliary elements localized upstream of the pA signal, are the upstream sequence elements (USEs). These U-rich sequences were originally identified in viral mRNAs. Like DSEs, USEs are poorly conserved but they were shown to modulate the recognition of pA signals. Both USE and DSE elements were shown to be present in the vicinity of all types of those pA signals that are used, canonical or weak variations, but they were essentially absent in the surrounding area of the silent pA sites (hexamers that are not used as pA signals). This suggests that the presence of a USE and/or DSE differentiates authentic pA sites from randomly occurring AAUAAA hexamers (19).

B. *Trans*-acting factors required for mRNA 3' end formation

The cleavage of the pre-mRNA and the addition of a pA tail require the recognition and the subsequent binding of the core C/P machinery to the *cis*-acting elements previously described. The core elements of the C/P machinery include many polypeptides and most of them included multi-subunits complexes (20) (Figure 2). The basal C/P machinery is composed by four different protein complexes and several single proteins. The Cleavage and Polyadenylation Specificity Factor (CPSF), that binds to the AAUAAA is composed by 5 subunits: CPSF-160, CPSF-100, CPSF-73, CPSF-30 and Fip1 (21); the Cleavage stimulation Factor (CstF) includes 3 subunits: CstF-77, CstF-64 and CstF-50; the Cleavage Factor I_m includes: CFI-68, CFI-59 and CFI-25 and the Cleavage Factor II_m containing Pcf11 and Clp1. Also part of the C/P machinery is the RNA Polymerase II (RNA pol II), Symplekin, Poly (A) Polymerase (PAP) and Poly (A) Binding Protein Nuclear 1 (PABPN 1).

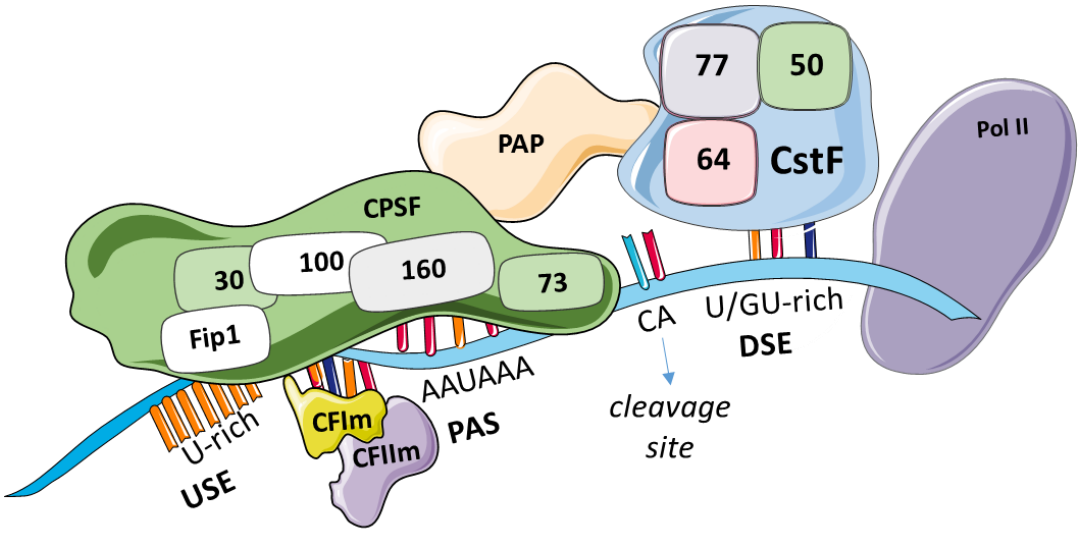


Figure 2. The cleavage and polyadenylation machinery involved in mRNA 3' end formation. The CPSF complex recognizes the pA signal; CPSF160 binds the AAUAAA signal and CPSF-73 executes endonucleolytic cleavage, preferably at CA dinucleotide. The CstF complex recognises the downstream sequence element (DSE) via the CstF-64 subunit. The numbers inside the boxes represent the members of each complex. Adapted from (20).

CPSF recognizes and binds to the pA signal hexamer through its CPSF-160 subunit (22). The CPSF-73 is the nuclease responsible for pre-mRNA cleavage (42). The CstF usually bind to the DSE *via* direct interaction of its CstF-64 subunit (23). CFI_m bind to UGUA motifs upstream of the cleavage site and stimulates binding of CPSF to pre-mRNA through a functional interaction with Fip1 (24, 25). Other proteins that also contribute for pre-mRNA 3' end formation are PAP that adds a pA tail to the newly formed 3' end (26-28) and the scaffold protein symplekin (29). PABPN1 binds to the pA tail and is involved in the export of the mRNA to the cytoplasm. Cleavage and polyadenylation are co-transcriptional reactions

that are connected to other processes mediated by RNA polymerase II (RNA pol II). The carboxy terminal domain (CTD) of RNA pol II acts as a platform to recruit important factors for this reaction, e.g. CstF (30), and thus it is considered to be part of the machinery (31).

2. USEs impact on mRNA 3'end formation regulation

A. The role of USEs in viral mRNAs

USEs were initially demonstrated to be U-rich sequences present upstream of the pA signal of many viral mRNAs and that were functionally interchangeable in different virus. Viruses (retroviruses, hepatitis B viruses (hepadnaviruses) and caulimoviruses) as well as many families of transposable elements produce unspliced, terminally redundant RNA molecules that serve as the template for reverse transcription and the pA signals are contained within each terminal repeat. Thus, for the synthesis of full length transcripts these signals must be ignored in the upstream portion (5' terminal repeat) of the RNA and recognized in the 3' terminal repeat. Viruses are capable of modifying the polyadenylation pattern of infected cells, sometimes in order to differentially express viral genes during the course of replication and in other cases as a way to supress the synthesis of cellular mRNAs in order to the viral ones to prevail (reviewed in(2)).

Simian Virus 40 (SV40) is a DNA virus with early and late phases of infection being differentially regulated. The 3' end formation was shown to be more efficient in the late than in the early pA signal, resulting in higher levels of steady-state mRNA. It was also shown that deletion of the sequences upstream of the late pA signal drastically decreased the pA site usage, identifying an USE (32). Therefore, the first identification of USEs came from the study of the late polyadenylation signal of SV40 also with the concomitant knowledge about the ability of this USE to confer the signal its exceptional strength.

The mechanism behind the differential pA site usage in the hepatitis B viruses was examined in 1990. The results disclosed multiple USEs, functionally analogous to other USEs existing in other retroid elements, which are required for correct pA site usage. (33).

In plants, experiments on cauliflower mosaic virus (CaMV) have shown that deletion of sequences upstream the pA signal result in a dramatically decreased processing at the CaMV pA site, also suggesting the presence of USEs with a positive role in the pA site usage (34).

The Adenovirus major late transcription unit codes for five collinear mRNA families, L1 to L5, each characterised by a unique pA site. In this case pA site usage is regulated according to the course of infection, with L1 being produced at early stages of infection and L2-L5 produced in late stages of infection. In 1989, the *cis*-acting elements located upstream of the pA signal were found to be responsible for L1 pA site choice (35).

Studies on the regulation of polyadenylation in the human immunodeficiency virus (HIV) showed that USE has a crucial role in enhancing the processing efficiency at the 3' long terminal repeats (LTR) (36-39). Brenton Graveley and Gregory Gilmartin showed in 1996 that the same USE increased the stability of binding of CPSF to the pA signal. The authors proposed this could be a general strategy for viral pA site selection (40).

B. The role of USEs in cellular mRNAs

USEs were also found to be present in cellular mRNAs. The first non-viral example was discovered in the C2 complement pre-mRNA (41). Although DSEs increase the efficiency of pA site specified by the adjacent hexamer, because they bind CstF, these elements are absent in the C2 complement pre-mRNA. Instead, the C2 pre-mRNA possesses an USE that activates mRNA 3' end formation, such as in the viral USEs (41). In 1998, Moreira *et al.* showed that the C2 complement USE enhanced both the cleavage and polyadenylation *in vitro* reactions and also that it could bind polypyrimidine tract-binding protein (PTB). Mutations on PTB binding site lead to a reduction in the efficiency of C2 mRNA 3' end formation both *in vivo* and *in vitro*. It was also shown that CstF-64 could also bind the C2 USE and that PTB and CstF compete for the binding to the USE. Therefore, a precise control of the levels of PTB and CstF in the cell are necessary for efficient usage of the C2 pA signal (42, 43).

One of the best known examples of regulated pA site selection is the immunoglobulin (Ig) M secretory, which occurs during B cell differentiation. It was proposed that regulation occurs by a complex set of factors, including competition with the splicing machinery, the presence of USEs and modulation in the binding of CstF to GU-rich DSEs required for polyadenylation to occur. When the USE was mutated, in the presence of a canonical pA signal 3' end formation still occurred, but when a non-canonical signal was used, there was a significant reduction in the efficiency of the mRNA 3' formation. Furthermore, USEs can retain a low level of polyadenylation activity in the absence of a functional pA signal sequence (44).

In 2002, highly conserved USEs were identified in three human collagen genes: COL1A1, COL1A2, and COL2A1. Apart from their structural role, collagen genes are important players in tissue development. By mutational analysis and cell transfections, USEs were shown to increase collagen polyadenylation efficiency both *in vitro* and *in vivo*. Moreover, it was also shown that the insertion of the USE motif of the SV40 pre-mRNA nearby a weak pA signal, such as the one present in adenovirus IVA_2, can enhance 3' end formation (45).

Cyclooxygenases (COX) are important enzymes in the production of prostaglandins. COX-1 is constitutively expressed while the activation of COX-2 expression is dependent on the stimulation through several signals. Bioinformatic analysis revealed that the human COX-2 gene has two different pA sites used in a tissue-dependent manner. In 2005, Hall-Pogar e*t al.* revealed that COX-2 USEs play a role in the regulation of 3'end formation. First, by *in vivo* polyadenylation assays in several cell lines, they demonstrated the presence of USEs of the proximal COX-2 pA signal. Secondly they showed that mutations in the USE sequence result in a significant decrease of the usage of this signal. Finally, they showed that the distance between USE and the pA signal is important for their function (46).

USEs were also found in the prothrombin (F2) mRNA. The main biological role of thrombin is in the regulation of the blood coagulation cascade. Essentially, it functions as a procoagulant as well as an anticoagulant, thus regulating homeostasis (47). Pro-thrombin pre-mRNA cleavage site is inefficient, however, a common mutation has revealed a gain-of-function of the 3' end processing mechanism. This mutation is sufficient to revert the physiologically inefficient pre-mRNA cleavage site into a most favourable CA cleavage site, resulting in an increase in the levels of pro-thrombin mRNA and protein. In 2004, Danckwardt *et al.* revealed that prothrombin mRNA 3' end is characterized by a weak downstream CstF binding site. In addition, they showed that there was an increase in the efficiency of the 3' end formation when U-rich elements were added upstream the pA signal. They also showed that despite the low-efficiency of the cleavage reaction and the CstF binding, the pA signal continued to be recognized because of the presence of U-rich USEs (47). Additionally, they identified the USE-binding proteins and characterised their action as *trans*-acting factors inducing pro-thrombin mRNA 3' end formation as U2AF35, U2AF65 and PTB, revealing another gene where the splicing and 3' end formation machineries interconnect in a USE-dependent manner (48).

USEs appear to have a significant role in genes with non-canonical pA signals, as these are weaker and therefore more prone to regulation (8, 44, 45). Auxiliary upstream U-rich elements are required in these signals to promote efficient mRNA 3' end formation.
Altogether, these findings lay emphasis on the importance of unravel all the players involved in an accurate mRNA 3'end formation. This is particularly important in genes with multiple pA signals, as regulation of pA site selection represents an important mechanism for spatial and temporal control of gene expression.

3. *polo* alternative polyadenylation

The mechanism through which more than one pA signal of the pre-mRNA becomes polyadenylated is named alternative polyadenylation (APA) (49). According to the pA signals position, APA can produce mRNAs that differ in their 3' untranslated region (UTR) length or can generate mRNAs encoding different proteins and ultimately no protein at all. Thus, APA is an important regulatory event in gene expression. Through alteration of the 3'UTRs length, APA can regulate processes such as mRNA stability, cellular localization and translation efficiency in certain developmental conditions (50). This is achieved by the binding of RNA-binding proteins (RBPs) to the 3'UTRs or targeting by microRNAs. RBPs may recognise consensus RNA sequence motifs, more degenerate sites and RNA secondary structures. Some of them act positively in gene regulation, for instance enhancing mRNA translation or stability while other RBPs can function as negative regulators by inducing specific mechanisms such as mRNA decay or translation inhibition. Different RBPs often compete for binding to the same mRNA sequence giving rise to dynamic post-transcriptional control mechanisms (42, 51, 52). APA in the 3' UTR diversifies the 3' UTR length that may be held in, or can escape from, regulation by RBPs or, eventually, silencing by micro-RNAs. The *polo* gene has two pA signals in the 3'UTR that are used to produce two mRNAs and this mechanism may be regulated by USE-binding proteins. *polo* codes for an important cell cycle kinase and it was the first described member of the Polo-like kinase family (53). This family includes extremely conserved serine/threonine proteins that share a common domain crucial for protein-protein interactions, the Polo-box domain (PBD) (54). Over the years, *polo* has been found to be involved in several biological processes: mitotic entry, spindle organization, mitotic progression and mitotic exit, rendering it a master regulator of the cell cycle (55). The two *polo* mRNA isoforms differ in the 3'UTR length due to the utilization of two pA signals by alternative polyadenylation (56). The transcripts also differ in the 5'UTR region, however, they have the same open reading frame and encode for the same protein, Polo. Both of the pA signals found in the 3'UTR are non-canonical, the first signal, AUUAAA, is the most common variation of the AAUAAA (the canonical signal) both in humans and *Drosophila melanogaster*. The other pA signal, AAUAUA is a more divergent variation of the canonical (being present only 5-10% of the *Drosophila* genes) (57) (Figure 3).

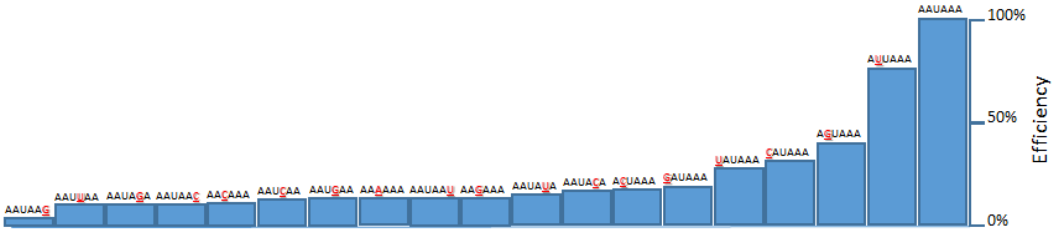


Figure 3. Efficiency of the pA signals variants. The AAUAAA and AUUAAA are the first and the second variants most efficient in both Human and *Drosophila.* Adapted from (57).

A. Physiological significance of *polo* alternative polyadenylation

Considering *polo* pA signals divergence from the canonical hexamer AAUAAA, the pA1 (proximal signal) is stronger than the pA2 (distal signal). Llamazares *et al*. showed that the 3' end formation occurs approximately 2-fold more efficiently in pA1 than in pA2 by Northern blot analyses (58). Therefore, it would be predictable that the shorter mRNA, the one using pA1, would have a more relevant function than the longer mRNA. If this would be the case, the longer mRNAs would be redundant and would not severely affect the amount of protein levels. To understand the biological role of *polo* pA signals, Pinto et al constructed transgenic flies producing only one *polo* mRNA isoform. Briefly, three constructs were produced using a previously described transgene *gfp-polo* as a template (59). The first includes the complete sequence of the 3'UTR with the two intact pA signals, the second included point mutations in pA1 (ATTAAA to GTTAAC), preventing it to be recognized and allowing only the expression of the longer isoform and the third transgene was made by deletion of pA2 and the entire downstream region, thus compromising the expression of the longer mRNA. To ensure that flies would express only one *polo* mRNA, the *polo*^{9} allele was used (60). polo^{9} was generated by P-element mediated mutagenesis and the P-element was inserted at position -34 upstream of the initiator ATG codon of the Polo protein. The insertion of this element results in a barely detectable amount of Polo protein levels. For this reason homozygous individuals for *polo*^{9} die at late third instar larval stage with the brains presenting a high mitotic index as a result of an arrest in pro-metaphase/metaphase stage of the cell cycle (60). Therefore, flies carrying *gfp-polo*, *gfp-polo*ΔpA1 or *gfp-polo*ΔpA2 transgenes in a *polo*^{9} /TM6B background were generated.

The analysis of the APA-derived phenotypes gave valuable insights on the importance of each *polo* transcript. The *gfp-polo*Δ*pA1*; *polo*^{9} flies present a very mild phenotype in the abdomen (Figure 4). In spite of the absence of the shorter isoform, most of the transgenic flies survived. But the absence of the longer *polo* in *gfp-polo*Δ*pA2*; *polo*^{9} flies leads to the dead of the flies with severe abdominal abnormalities (Figure 4). The phenotype

is characterized by the absence or incorrect formation of tergites. This is an evidence of the importance of the longer *polo* transcript in the correct formation of the adult abdomen (56).

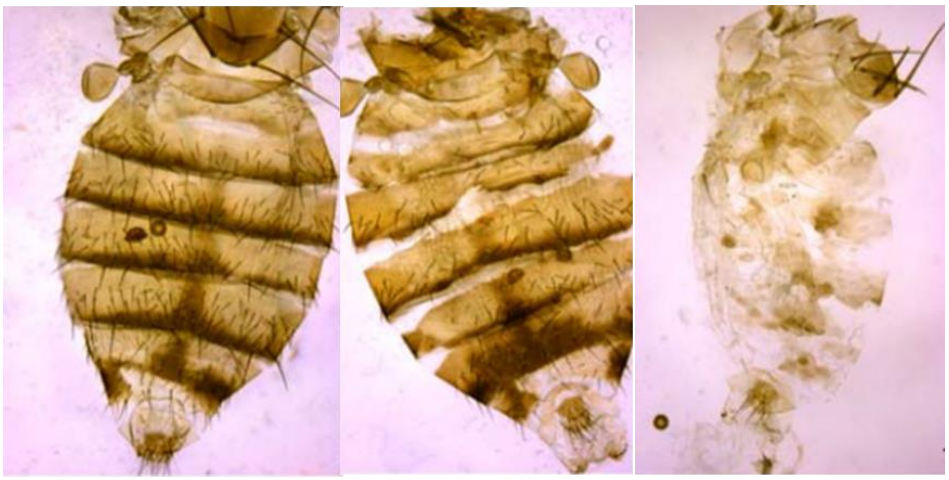


Figure 4. Levels of the longer *polo* mRNA isoform are critical to the correct formation of the adult abdomen. Adult abdomen of *gfp-polo, gfp-polo*Δ*pA1* and *gfp-polo*Δ*pA2* flies in a *polo*^{9} homozygous background is depicted in the figure. Adapted from (56).

By luciferase assays, Pinto et al. showed that pA2 is responsible for more than 50% of the Polo levels (56).As Polo is a limiting factor in the cell cycle, the most plausible explanation for the *gfp-polo*Δ*pA2* flies' phenotype was that Polo protein produced by using the shorter isoform is not sufficient to ensure the rapid proliferation of histoblasts at the onset of metamorphosis. Since proliferation was disrupted it leads to an inaccurate development of the adult epidermis resulting in abdominal malformation, and more importantly, lethality (56).

Other experiments using GAL4/UAS showed that an overexpression of Polo results in a switch in the pA site selection, from pA2 to pA1 signal recognition. Therefore, a down-regulation of Polo expression occurs since pA1 produces a functional transcript but that is not efficiently translated. Such results suggest an autoregulatory feedback loop mechanism (Figure 5) for this gene (56).

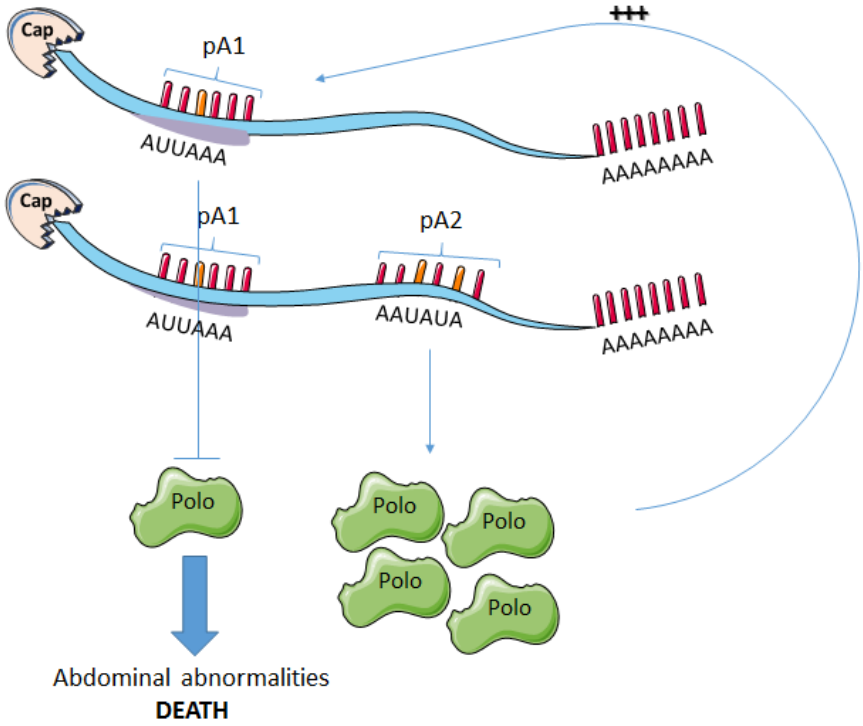


Figure 5. *polo* autoregulatory feedback loop mechanism. When Polo is overexpressed, pA1 signal is predominantly selected, leading to a decrease in the levels of Polo protein produced, which are critical to organism development. Adapted from (56).

B. USE1 – an highly conserved element in *polo* 3' UTR

To understand what regulates the switch in the pA signal used, the 3'UTR was searched for conserved sequence motifs (61). Using *in silico* analysis, the genomes from twelve *Drosophila* species and also mosquito (*A. gambiae*), honeybee (*A. mellifera*) and the beetle (*T. castaneum*) were aligned and using Phastcons three different conserved sequences were found:USE1, USE2 and USE3. The first sequence was considered the best candidate for further investigation because, although it is not the most conserved, it is the most T-rich element and resembled the USEs already described for other genes.

To study the physiological function of the USE1, flies with a deletion in USE1 (*gfp-polo∆USE1; polo*^{9}) were generated. The phenotype for both *∆USE1* and *∆pA1* transgenic flies is very similar (Figure 6) showing that a deletion of pA1 signal has the same physiological result that the deletion of the regulatory element USE. This suggests that USE1 functions in pA1 site usage.

Figure 6. *gfp-poloΔpA1;polo*^{9} flies have an abdominal phenotype identical to *gfp-poloΔUSE1;polo*^{9} *flies*. Adult abdomen of *gfp-polo, gfp-poloΔpA1*and *gfp-poloΔUSE1* flies in a *polo*^{9} homozygous background is depicted in the figure. Adapted from (61).

With the purpose of understanding how USEs could regulate *polo* pA site choice, Pinto *et. al*. searched for proteins that could bind these sequences, using U.V. crosslinking assays. Heterogeneous nuclear ribonucleoprotein (hnRNP) C and PTB were found to bind USE1 and USE3. HuR was the only protein that was found to bind specifically to USE1 and does not bind upstream of the second pA site while the other two proteins are also capable of binding upstream the pA2 (Figure 7).

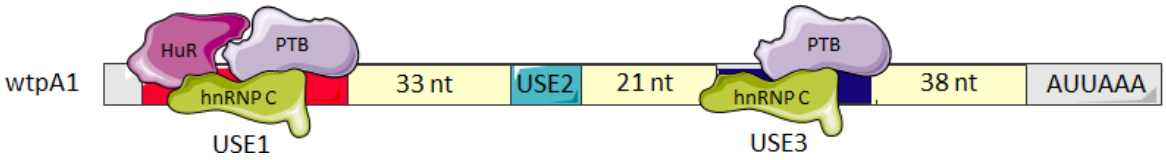


Figure 7. Schematic representation of the binding sites for HuR, hnRNP C and PTB. HuR binds specifically to the USE1 and PTB and hnRNP C bind to the USE1 and USE3. Adapted from (61).

C. USE-binding proteins – are they modulators of *polo* pA site usage?

a) HuR/Elav

Elav (embryonic lethal abnormal vision) was originally identified in *Drosophila*. Later on, Elav structural homologues have been identified in a wide range of organisms, including humans, suggesting an important conserved role for the protein (62). The name *embryonic lethal abnormal vision* emerged from the finding that, in *Drosophila melanogaster*, the lack of this gene leads to embryonic lethality in null mutants and to deficiencies in the visual system in hypomorphic alleles. In mammals ELAV-like 1 gene (HuR) is ubiquitously expressed while the other members of the family for instance ELAVL2 (HuB), ELAVL3 (HuC), and ELAVL4 (HuD) are the neuron-specific members of this family.

elav encodes a neuronal nuclear RBP with a specific role in the alternative splicing of some target mRNAs (63, 64) inclusive its own mRNA (65) and alternative polyadenylation (66, 67). The Elav family in *Drosophila* contains two additional members expressed in neurons, Fne (found in neurons) and Rbp9 (RNA-binding protein 9). Rbp9 is also expressed in gonads. The family of Elav/Hu proteins is important at many steps of post-transcriptional regulation of gene expression, from nuclear events, such as alternative splicing and alternative polyadenylation, to cytoplasmic events such as the regulation of mRNA stability, localization and translation. Past studies showed that this protein is required for alternative splicing of *armadillo* (*arm*), *neuroglian* (*nrg*), and *erect wings* (*ewg*), stimulating the production of neural-specific isoforms (68, 69). Interestingly, Elav has been found to function by suppressing proximal pA signals, thereby allowing read through of elongating polymerase II complexes. The data reveals another nuclear role of *elav* in the fly which is the inhibition of mRNA 3'-end processing by binding at the vicinity of a pA site (69). Furthermore, it was also shown that ectopic expression of Elav in non-neural tissues induces ectopic expression of longer mRNA isoforms. This was the first evidence that Elav coordinates the expression of 3' UTR extensions during neural development (66). More importantly, Zhu *et al.* using *in vitro* polyadenylation assay had previously shown that Hu proteins selectively block both cleavage and polyadenylation at sites containing U-rich sequences similarly to the previously described USEs. Moreover, Hu proteins do not have an effect on pA sites that do not contain U-rich sequences or sites in which the U-rich sequences are mutated (70). Recently, Rogulja-Ortmann *et al.* identified Elav as a key regulator of *Ultrabithorax* (*Ubx*) RNA processing, expression and function within the *Drosophila* embryonic central nervous system (CNS). They also showed that Elav is able to modify the alternative polyadenylation pattern (67). Given that Elav regulates alternative polyadenylation of different genes, that the majority of pre-mRNAs are alternatively polyadenylated and that USEs are regulators of polyadenylation, then, understanding how Elav/Hu proteins regulate alternative polyadenylation is crucial. Altogether, these findings suggest that due to the Elav involvement in several steps of mRNA processing, all the way from the nucleus to the cytoplasm, it is likely that it makes an extensive spectrum of interactions with several cellular factors, especially with other RBPs.

b) PTB/Heph

Another RNA-binding protein that binds to *polo* USE1 is hnRNP I/PTB. hnRNPs proteins bind pre-mRNA molecules in the nucleus and promote the correct mRNA processing, nuclear export and metabolism. PTB has the ability to shuttle between the nucleus and cytoplasm (71). In the nucleus, PTB acts as an auxiliary factor in alternative

splicing (72-80) and polyadenylation (42, 43, 48, 76) while in the cytoplasm it is involved in post-transcriptional regulation processes such as translation (81, 82). One of the roles of PTB in 3'end processing was elucidated when it was shown that PTB enables the hnRNP H binding to upstream and downstream auxiliary elements which in turn stimulates the assembly of 3' end processing factors (PAP and CstF), promoting both cleavage and polyadenylation (83).

In *Drosophila*, *Hephaestus (heph)* encodes the homologue of mammalian PTB (84). *Drosophila* PTB was also shown to be involved in *oskar* mRNA processing. It binds to multiple sites in the 3'UTR of *oskar* pre-mRNA inhibiting its translation while unlocalized, and activating its translation when the mRNA arrives to the posterior pole of the oocyte (85). In the model suggested PTB stimulates the formation of compacted RNP particles with *oskar* RNAs, thus inhibiting the access of the translation machinery. It still remains to be shown if *Drosophila* PTB is able to regulate alternative polyadenylation.

c) hnRNP C/CG42458

The third RNA-binding protein that was found to bind to *polo* USE1 is hnRNP C. The hnRNP C protein exhibits two different isoforms, hnRNP C1 and hnRNP C2 (86) that present multiple functions in the eukaryotic cells. The binding of hnRNP C1/C2 to 3′UTRs and subsequent stabilization of specific mRNAs has been well documented (87-90) and they were also found to have a role in splicing (91).

In 1990, it was found that hnRNP C is likely to bind to the UUUUU sequence and that this element can restore the efficiency of 3' end formation of three independent pA sites (92). Christian *et al.* showed that hnRNP C1/C2 binds p53 mRNA and regulates p53 expression *via* specific binding to regulatory *cis*-element identified in the 5' coding region of the gene (93).

Although it is clear that APA is a wide-spread mechanism for the control of gene expression that ultimately result in significant biological consequences, the mechanisms underlying pA site selection, in particular in *Drosophila*, still remain to be clarified. Investigating these mechanism and the role of RBPs in such regulation could provide valuable insight about the switch in the polyadenylation pattern and consequently in gene expression in specific tissues.

4. Aims

Previous results show that the *cis* sequence element USE1, localized upstream of *polo* pA1, binds HuR/ELAV and PTB/Heph. However, their role in *polo* mRNA isoforms expression regulation was not investigated so far. Therefore the aims of this thesis were to determine if the USE1 sequence has a role in *polo* pA site usage and to elucidate the function of these *trans*-acting factors in *polo* mRNA 3' end formation and APA..

II. Material and methods

1. *Drosophila* stocks

Transgenic flies w^{1118}; *gfp-polo; polo*^{9}*/TM6B*, w^{1118}; *gfp-polo*Δ*pA1; polo*^{9}*/TM6B* and w^{1118}; *gfp-polo*Δ*USE 1; polo*^{9}*/TM6B* used in this study have been previously characterized (61). Adult flies were selected regarding their gender and genotype and only males homozygous for *polo*^{9} in the third chromosome were collected for further studies. The phenotypical trait that differs between the homozygous or heterozygous is based on the presence of TM6B. TM6B is a balancer chromosome and Humeral (*Hu) allele* is the dominant marker. Homozygous flies present the wild-type (WT) phenotype that corresponds to the presence of two bristles. Heterozygous flies present additional Hu bristles with some being slightly shorter than normal. The heads and bodies of the homozygous flies were collected separately, frozen in 0.2 ml of TRIzol (Invitrogen) and stored at -80ºC.

During the course of the experiments flies were maintained at 25ºC using standard culture conditions and transferred to new vials with fresh medium every 3 days. Fly culture medium was composed by mixture A (2000 ml distilled deionized (ddH_2O) + 103 ml Honey + 24 g Agar) and mixture B (1000 ml ddH_2O+240g Cornmeal + 54 g Yeast extract + 30 g Flour of soy + 60 g Malte). Mixture B was added to the previously boiled mixture A and heat until boiling for 35 min. The final mixture was allowed to reach at least 60ºC before adding 17.4 ml of propionic acid (85% (v/v)).

2. Cell culture and freezing

Schneider's *Drosophila* line 2 (S2 cell line) was derived from a primary culture of late stage (20-24 hours old) *Drosophila melanogaster* embryos. Some features of this cell line suggest that it is derived from a macrophage-like lineage. Cells were kept in 25 cm^2 culture flasks without ventilation (Starstedt) in a cell culture incubator at 25ºC with no CO_2. Cells were routinely sub-cultured every 3-4 days in Schneider's Insect medium (Sigma) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco). Cells were detached from the culture surface by gently pipetting up and down and reseeded in new cell culture flasks at a 1:5 ratio of cell suspension to fresh medium. Cells were not grown continuously for over six months in order to avoid the accumulation of small changes. For that, several stocks of cells were frozen in order to be used when necessary. Cells were frozen as follows: first they were grown to subconfluence of approximately 1-2 x 10^7 cells/ml.

Secondly, they were removed from flask and centrifuged at 1200 rpm 5 min. The medium was discarded and cells were resuspended at approximately 1.1 x 10^7 cells/ml in freezing medium (FBS + 10% DMSO, filtered). Finally, 1ml cell suspension was aliquot per cryovial and stored.

3. Single fly genomic DNA extraction

Each fly was collected and placed in a 0.5 microcentrifuge tube. The squishing buffer was prepared according to the following: 10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl and Proteinase K was added to a final concentration of 200 µg/ml. Fresh Proteinase K was added each time. Flies were mashed for 5 - 10 seconds with a pipette tip containing 50 µl of squishing buffer and incubated at room temperature for 20-30 minutes. Proteinase K was inactivated by heating to 95 °C for 1-2 minutes. 1µl of DNA was then used to the PCR reaction.

4. Total RNA isolation

Total RNA from cells/tissues was isolated using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions, but with some modifications for fly tissues. Twenty flies (frozen as mentioned above) were homogenized with a disposable plastic pestle and 0.8 ml of TRIzol was added to the mixture. Samples were incubated at room temperature for 5 min and insoluble debris such as exoskeleton was removed by centrifugation at 4ºC (12.000 rcf, 10 min). Supernatant was collected into new microcentrifuge tubes, chloroform (200µl) was added, and tubes were vigorously shaken by hand for 15 sec. After a 3 min incubation at room temperature the mixture was centrifuged (10.000 relative centrifugal force (rcf)) for 15 min at 4ºC. The upper aqueous phase where RNA remains was then transferred to a new RNase-free microcentrifuge tubes. Total RNA was precipitated by mixing with 500µl of isopropanol and incubated at room temperature for 10 min. Tubes containing the precipitated RNA were centrifuged at 12.000 rcf for 10 min (4ºC), the supernatant was removed and the pellet was washed with 1 ml of 70% ethanol. The supernatant was once again removed after a 5 min centrifugation at 5.000 rcf at 4ºC and the RNA pellets were allowed to dry for 10 min at room temperature. Finally, the isolated RNA was resuspended in 30 µl of RNase-free water previously heated to 55ºC.

5. Quantity and Quality analysis of total RNA

Total RNA was quantified by measuring the absorbance of each sample at 260 nm (A_{260}) in a NanoDrop 1000 spectrophotometer (Thermo Scientific). One unit of A_{260} corresponds to 40µg/ml of RNA (94). Additional measurements were made at 230 nm and

280 nm to assess the purity of each RNA sample. Nucleic acids absorb at 260 nm, proteins mostly at 280 nm and organic compounds (such as phenol) at 230nm, thus RNA purity was considered satisfactory when A_{260}/A_{280} ratio was above 1.9 and the A_{260}/A_{230} ratio was in the 2.0 - 2.2 range.

6. DNase treatment

Prior to reverse transcription polymerase chain reaction (RT-PCR) samples were treated with RNase-free DNase I (Roche) for 30 min at 37ºC: 12 U of DNase I, 1x Incubation buffer, 1µg of total RNA and RNase-free water up to 12 µl. To stop the reaction, 8 mM of EDTA (pH=8) was added and the sample was heated to 75ºC for 10 min. The RNA integrity was analysed using an Experion RNA StdSens Analysis Kit (Bio-Rad) with and Experion automated electrophoresis station. The results indicated that the RNA was intact after DNase treatment.

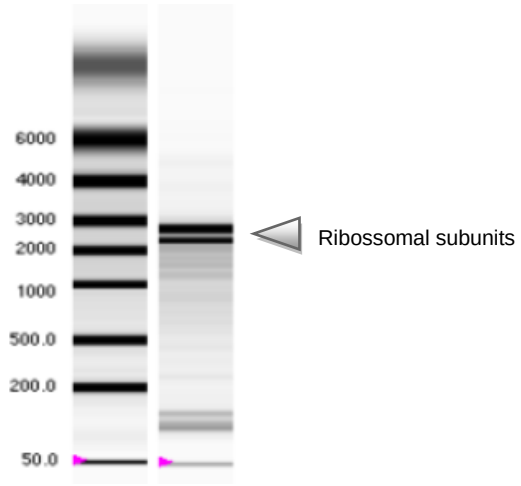


Figure 8. Electropherograms run with Bio-Rad Experion RNA Integrity of RNA samples after the DNase treatment. The ladder is presented on the left and the RNA sample on the right. Drosophila 28S rRNA is processed into 2 fragments that migrate in a similar manner to the 18S rRNA.

By running RT minus control, it was possible to confirm that there was no genomic DNA detected by real-time polymerase chain reaction (qPCR) after the DNase I treatment.

7. cDNA synthesis

For complementary DNA (cDNA) synthesis, random hexamers (2.5 µM, Sigma-Aldrich), deoxyribonucleotides (dNTPs) (0.5 mM, Thermo Scientific) and 1µg of RNA were mixed and incubated for 5 min at 65 ºC on a 48-well TPersonal Thermocycler (Biometra). Followed by 5 min on ice, 1x First-Strand Buffer (Invitrogen), 5mM dithiothreitol (DTT) (Invitrogen), 1 U/µL RiboLock RNase Inhibitor (Thermo Scientific) and 5 U/µL SuperScript III

Reverse Transcriptase (Invitrogen) were added to the reaction and gently mixed. Samples were then incubated for 5 min at 25ºC, 60 min at 50ºC and 15 min at 70 ºC.

8. Oligonucleotides design

All oligonucleotides (Sigma-Aldrich Quimica SL) were designed with Primer3 (95, 96). Oligonucleotides used in PCR and qPCR quantifications are listed in Table 1 and oligonucleotides used to amplify the fragments for *in vitro* transcriptions are listed in Table 2. Oligonucleotides for *in vitro* transcription were designed according to the cDNA or exon sequences of the target gene, preceded at the 5' end by the T7 promoter sequence (TAATACGACTCACTATAGGGA, the transcription start site for T7 RNA polymerase is underlined). A BLAST analysis was done to ensure that oligonucleotides do not contain complete 19-mer homology to other *Drosophila* genes and thus, ensure that dsRNA is specific for the target gene (97).

9. Real-time quantitative PCR (qPCR)

For gene expression quantifications by RT-qPCR the efficiency of the qPCR oligonucleotides was assessed by a cDNA serial dilution (1:10) standard curve. The cDNA dilutions included ranged from the lowest to the highest expression levels of each target being quantified. Plotting the dilution factor against the threshold cycle (C_T), obtained for each dilution, results a linear regression from which the amplification efficiency (E) can be calculated:

$$E = 10^{\left(-\frac{1}{slope}\right)}$$

Considering that a given PCR amplicon doubles its amount during the geometric phase of the amplification, then a 100% efficient PCR reaction can be characterized by 2^n and therefore,

$$2 = 10^{\left(-\frac{1}{slope}\right)}$$

This implies that the slope of the cDNA serial dilution (1:10) standard curve should be -3.32. As such, different annealing temperatures and oligonucleotide concentrations were tested in order to obtain the highest possible efficiency (90-110%). Each qPCR reaction contained 1 µl of cDNA, 0.5x of SYBR Select Master Mix (Invitrogen), variable concentrations of each oligonucleotide pairs (Table 1) and ddH_2O up to 20µl.

Reactions were performed in a StepOne Real-time PCR System thermocycler (Applied Biosystems) with the following program: 2 min at 50ºC; 2 min at 95ºC, 40 cycles of 15 sec at 95 °C, 15 sec at the respective annealing temperature and 1 min at 72ºC.The housekeeping genes tested to standardize the amount of sample added to each reaction were the Ribosomal protein L32 (*rp49,* FBgn0002626) and the Signal recognition particle (*7SL*, FBgn0000003).

Gene expression quantifications were performed by the ΔC_T method. This method assumes maximal amplification efficiencies (90-110%) for both target and reference genes to enable the relative quantification of the target gene in the different samples. The following equation was therefore used:

$$2^{\Delta C_T} = 2^{C_{T(reference)} - C_{T(target)}}$$

RNAi experiments were analysed using the $2^{-\Delta\Delta Ct}$ method, which uses the same mathematical assumptions. This method directly compares the target gene expression level on a test (treatment) sample relatively to a calibrator (control):

$$\Delta\Delta C_T = \Delta C_{T(test)} - \Delta C_{T(calibrator)}$$

in which,

$$\Delta C_{T(test)} = C_{T(target,test)} - C_{T(reference,test)}$$

$$\Delta C_{T(calibrator)} = C_{T(target,calibrator)} - C_{T(reference,calibrator)}$$

Since the qPCR oligonucleotides were previously designed and optimized for maximal efficiency $\mathrm{E_{target}} = \mathrm{E_{reference}} = 2$, the fold-change of the target gene in the test sample relative to the calibrator sample is

$$\text{Normalized fold expression (Fold induction)} = \frac{2^{\Delta T_C, target\,(calibrator-test)}}{2^{\Delta T_C, ref\,(calibrator-test)}} =$$

$$= 2^{-\left[\left(C_{T(target,test)}\right) - C_{T(target,calibrator)}\right] - \left[\left(C_{T(reference,test)} - C_{T(reference,calibrator)}\right)\right]}$$

$$= 2^{-\Delta\Delta C_T}$$

Table 1. Sequences and oligonucleotide conditions used for PCR and qPCR amplification: amplified product length (bp), annealing temperature (ºC) and the concentration (µM) used.

PCR and Real-time PCR oligonucleotides					
Target gene	**Oligonucleotide forward 5'->3'**	**Oligonucleotide reverse 5'->3'**	**Product length (base pairs)**	**Annealing temperature (ºC)**	**Concentration (µM)**
***polo* pA1+pA2**	CCGTACAACATGTGCCGTAG	CTTTAGACACGCCGTTCTCC	176	58	0.15
***polo* pA2**	ACGTGTTTCGAAATGCCTAT	ACACTTAAACACTTTGCAGCAG	162	58	0.5
cg42458	ACGTCAAGCTGCTGAATGG	AAGGTACGGACAATGCAAGC	100	58	0.3
heph	ATCACACGTATCGGCTTTCC	CACAGCCATGTCTCACTT	57	58	0.15
elav	GGCTTTGTTGGTCTTGAAGC	AGGATCCCACAACGAATCAG	143	58	0.15
rp49	ATCGGTTACGGATCGAACAA	GACAATCTCCTTGCGCTTCT	165	58	0.15
7sl	TTGGCTAAGGAGGGATGAAC	CTACTGCCTACCACGGGAAC	70	58	0.15
P-element F	ACGCTTCACGTCCAATTCAT	GCTATCGACGGGACCACCT	276	56	0.5
P-element R	ACGCTTCACGTCCAATTCAT	GCTGCCCTGGTGGAATACAA	271	56	0.5

10. PCR and electrophoresis

Polymerase chain reaction (PCR) was carried out with 1U GoTaq Flexi DNA Polymerase (Promega), 1x Green GoTaq Reaction Buffer (Promega), 5 mM $MgCl_2$ (Promega), 0.5 mM dNTPs (Thermo Scientific), 0.5 µM of each oligonucleotide primer, 2.5 ng/µl of cDNA and ddH_2O up to 50µl. Reactions to obtain DNA templates for *in vitro* transcription were performed in a 48-well TPersonal Thermocycler (Biometra) with the program described in Table 2.

All the PCRs were performed using cDNA of S2 cells as template except the Red fluorescent protein (*DsRed*) DNA was obtained by PCR on the vector pIRES2 DsRed-EGFP kindly donated by Prof. Bin Tian using conditions described in Table 2.

PCR products were analysed by gel electrophoresis on a 1.5% TAE-agarose gel stained with SybrSafe DNA gel stain (Invitrogen) and using the GeneRuler DNA Ladder Mix (Thermo Scientific) under the UV light of a Gel Doc XR+ system (Bio-Rad).

11. PCR products purification

Purification of the PCR products was performed using the QIAquick PCR Purification Kit according to the manufacturer's instruction (Qiagen). The concentration of the products was measured using NanoDrop 1000 Spectrophotometer (Thermo Scientific) and the quality of the DNA was analysed on an agarose gel.

Table 2. Sequences and oligonucleotide conditions used to amplify DNA templates for *in vitro* transcriptions. T7 promoter sequence is represented by the small letters and the sequence of the oligonucleotide that is specific for target gene amplification is in capital letters. Amplified product length (bp), and thermal cycler program are also described.

Sequences and oligonucleotide conditions used to amplify DNA templates for *in vitro* transcriptions

Target gene	Oligonucleotide forward 5'->3'	Oligonucleotide reverse 5'->3'	Product length (base pairs)	Thermo cycle program
dsRed	gaattaatacgactcactatagggaga CTTCAAGGTGCGCATGGAG	aattaatacgactcactatagggaga GGACTTGAACTCCACCAGGTAGTG	552	95ºC-5min 29 cycles, (95ºC-30sec; 60ºC-30sec;72ºC-30sec) 72ºC-7min
elav	taatacgactcactatagggaga GGCAATGATAGCCCTTGTG	taatacgactcactatagggaga CAATACGAATGGCAATGCAG	594	95ºC-5min 3 cycles (95ºC-30sec; 56ºC-30sec; 72ºC-1min) 5 cycles (95ºC-30sec; 61,5ºC-30sec; 72ºC-1min) 27 cycles (95ºC-30sec; 65ºC-30sec; 72ºC-1min) 72ºC-7min
heph	taatacgactcactatagggaga ACATTCAGAGCCGTCAGCTT	taatacgactcactatagggaga CATAGCGACTACAGCGTCCA	477	95ºC-5min 3 cycles (95ºC-30sec; 58ºC-30sec; 72ºC-1min) 5 cycles (95ºC-30sec; 61,5ºC-30sec; 72ºC-1min) 27 cycles (95ºC -30sec; 65ºC-30 sec; 72ºC-1min) 72ºC-7min
cg42458	taatacgactcactatagggaga CTTGGGATTTGATGGTCCAC	taatacgactcactatagggaga TTCAGTTCGTCGGAGTGATG	587	95ºC-5min, 3 cycles (95ºC -30sec; 56ºC-30 sec; 72ºC -1min) 5 cycles (95ºC-30sec; 58ºC-30 sec; 72ºC -1min) 27 cycles (95ºC-30sec; 60ºC-30 sec; 72ºC -1min) 72ºC-7min

12. dsRNA synthesis

Double-stranded RNA was generated by *in vitro* transcription using the MEGAscript T7 Kit (Ambion). Briefly, DNA with opposing T7 promoters at the 5'ends of each strand was used as template and a single transcription reaction was performed. Each reaction contained 1 µg of DNA template, 1x T7 Enzyme Mix, 1x Reaction Buffer, 7.5 mM of each ribonucleotide and RNase-free water up to 25µl. The reaction components were added according to manufacturer's instruction, mixed gently and incubated at 37°C for 16 hours. The dsRNAs were purified with the PureLink RNA Mini Kit (Ambion) following the manufacturer's instructions for the purification of RNA from liquid samples. Succinctly, to one volume of liquid sample, one volume of lysis buffer with β-mercaptoethanol and one volume of 96-100% ethanol were added. The reagents were mixed by pipetting up and down and the mixture was transferred to a spin cartridge and centrifuged at 12.000 × g for 15 sec at room temperature. Wash Buffer II with ethanol (500 µl) was added to the each spin cartridge and centrifuged at 12.000 × g for 15 seconds at room temperature. The spin cartridge was once again centrifuged at 12.000 × g for 1 minute at room temperature to dry the RNA-bound membrane. Each spin cartridge was left at room temperature for 1min after adding RNase-free water (30µl) to its centre and centrifuged for 2 min at 12.000 × g (room temperature) to elluted the RNA into the recovery tube.

Finally, to anneal the newly *in vitro* transcribed RNA strands, the reaction mix was incubated in a 48-well TPersonal Thermocycler (Biometra) with the following program: 96ºC for 5 min and 1 min for each further cycle decreasing the temperature by 1ºC sequentially until 24ºC. The concentration of dsRNA was quantified by measuring the absorbance of a 1:10 dilution of the sample at A_{260}. The quality of the dsRNA was evaluated on 1% agarose gel.

13. RNAi in *Drosophila* S2 cells

Drosophila S2 cells were sub-cultured to $1x10^6$ cells/ml in 5 ml of regular growth medium and allowed to proliferate for 48 hours. Cells were then homogenised and counted (1:1 mixture of cell suspension and 0.4% Trypan blue (Invitrogen)) on a TC10 automated cell counter (Bio-Rad). If the cell number corresponded to an exponential $4x10^6$cells/ml, cells were centrifuged and resuspended in culture medium without FBS to a final concentration of $1x10^6$ cells/ml. The cell suspension was distributed to each well of a 6-well plate ($1x10^6$ cells/well) and the dsRNAs were added drop-wise while mixing by swirling. After 60 min at 25ºC, 2 ml of regular growth medium was added to each well. The concentration of each

dsRNA and respective incubation time was optimized for every target. dsRNA for *DsRed,* which codes for red fluorescent protein of *Discosoma sp*., was used as a control to distinguish sequence-specific silencing from non-specific effects in the RNAi experiment, since it does not target any known mRNA in the cell.

14. Statistical analysis

The significance of differences was determined by two tailed Student's t-test analysis with 95% confidence intervals, calculated with GraphPad Prism 6 (GraphPad Software). Differences were considered to be statistically significant if $p<0.05$.

III. Results

1. USE1 affects *polo* alternative polyadenylation pattern

A. Model establishment

The USE1 of *polo* was found to influence polyadenylation reaction *in vitro* (unpublished data) but its function *in vivo* had yet to be identified. For that purpose, we used different transgenic lines *w^{1118};gfp-polo;polo^{9}/TM6B* (*gfp-polo*), *w^{1118};gfp-poloΔpA1;polo^{9}/TM6B* (Δ*pA1*) and *w^{1118};gfp-poloΔUSE1;polo^{9}/TM6B* (*ΔUSE1*) already available in the laboratory. To confirm and detect the presence of the transgene in the fly strains used, we performed a PCR using specific oligonucleotides to amplify a fragment of *gfp* in all the transgenic lines. Figure 9A shows that it is possible to amplify *gfp* in *gfp-polo* transgenic line, confirming that the pW8-*gfp-polo* transgene was inserted in the genome. This analysis was also performed in all the other transgenic flies (data not shown) and it was observed that all the fly stocks carry the specific transgene. Flies carrying the transgenes were generated in a *polo^{9}/TM6B* background. *polo^{9}* contains a P-element inserted downstream of the second transcription start site of *polo* and it is the strongest hypomorphic allele described to date. As a consequence, both *polo* mRNAs production is affected and homozygous individuals for *polo^{9}* die at late third instar larval stage. Pinto *et al.* determined that *polo* mRNA and subsequently protein levels contributed by the transgenes were sufficient to rescue the third instar larval lethal phenotype exhibited by *polo^{9}* individuals in the transgenic lines *w^{1118};gfp-polo;polo^{9}*, *w^{1118};gfp-poloΔpA1;polo^{9}* and *w^{1118};gfp-poloΔUSE1;polo^{9}*. Moreover, Pinto *et al.* established that the transgenes were responsible for almost all *polo* mRNA levels in all the transgenic lines and that endogenous *polo* mRNA contribution was insignificant (56). To verify that all the transgenic lines still have the same features they had when created, oligonucleotides in the *polo* 5' UTR and inside the P-element were designed and used (Fig. 9B) to detect the presence of the P-element as well as the direction in which it was inserted. Figure 9B shows that the transgenic lines carrying pW8-*gfp-polo* also have the *polo^{9}* mutant allele. The same analysis was done in the other transgenic flies and it was possible to amplify the fragments in all of them (data not shown).

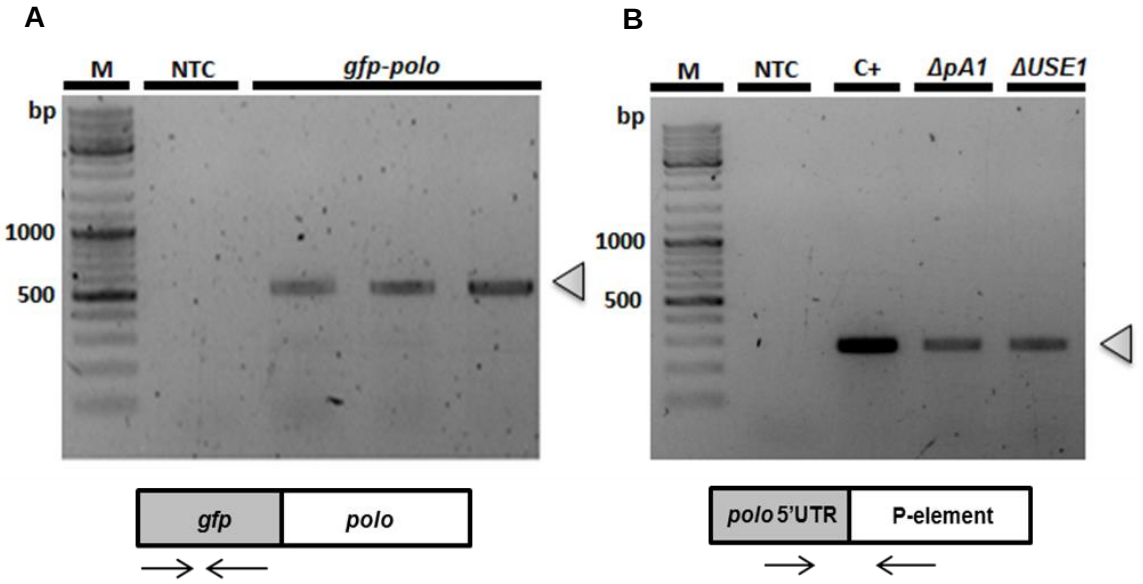


Figure 9. Transgenic flies contain the transgene and the polo[9] allele. Whole bodies of adult *gfp-polo* flies were collected and the genomic DNA of single flies was extracted to perform the PCR reactions. Oligonucleotide pairs used to amplify a fragment of *gfp* and a portion between *polo* 5'UTR and the beginning of the P-element were design as depicted in the figure. The arrows indicate the bands that confirm the presence of *gfp* and polo[9], respectively. M - DNA ladder; NTC – No template control; C+-positive control *gfp-polo*, *ΔpA1* and *ΔUSE*1flies.

At this stage, it was possible to conclude that our model was suitable to study the influence of the USE1 in *polo* alternative polyadenylation.

B. Real-time PCR (qPCR) optimization

All the oligonucleotides had to be optimized to reliably quantify *polo* mRNA expression levels by qPCR. This included optimization of the oligonucleotide pair to measure total *polo* mRNA levels (pA1+pA2), the oligonucleotide pair to measure only the longer isoform (pA2) and the oligonucleotide pair to amplify an endogenous control. Annealing temperature was first optimized for each oligonucleotide pair by doing a gradient PCR. Secondly, a standard curve was performed to each primer pair using cDNA serial dilution (1:10), and the annealing temperature was adjusted using qPCR when necessary. The cDNA dilutions ranged from the lowest to the highest expression levels of each target (see Material and Methods). Under specific conditions (see Table 1), all the oligonucleotide pairs tested had an efficiency of 2, *i.e*. 100% efficiency, which allowed using the ΔC_T method and $2^{-\Delta\Delta Ct}$ method (see Material and Methods) in the subsequent data analysis of the quantification of gene expression by RT-qPCR.

Before quantifying *polo* mRNA isoforms, it was necessary to select a gene that could be used as an endogenous control. Ribosomal protein L32 (*rp49)* and Signal recognition particle (*7SL*) are housekeeping genes described in the literature as commonly used in

Drosophila. Thus, we measured the C_T values of *rp49* and *7SL* mRNA expression in all of the transgenic lines. Although the C_T values of *7SL* mRNA expression differed among the transgenic lines, the C_T values of *rp49* mRNA expression were similar in all of the transgenic lines. Therefore, ΔC_T ($C_{T(target)}$-$C_{T(control)}$) was close to zero in the housekeeping gene *rp49* (Table 3) meaning that its expression is not affected by the experimental treatment under study and it can be used as an endogenous control of the experiment.

Table 3. Quantification of *rp49* mRNA levels in male bodies of two different transgenic lines and comparison with the control, *w^{1118}; gfp-polo; $polo^9$*.

Target (transgenic flies)	ΔC_T ($C_{T(target)}$-$C_{T(rp49)}$)
w^{1118};gfp-poloΔpA1;$polo^9$	0.315
w^{1118};gfp-poloΔUSE1;$polo^9$	0.338

C. Quantification of *polo* mRNA isoforms expression levels by RT-qPCR

polo mRNA isoforms expression levels were quantified by RT-qPCR using the primer pairs whose optimization was described in the previous section. Quantifications were done in male bodies of *w^{1118}; gfp-polo; $polo^9$* flies, *w^{1118}; gfp-poloΔpA1; $polo^9$* flies and *w^{1118}; gfp-poloΔUSE1; $polo^9$* flies (herein referred as *gfp-polo*, *ΔpA1* and *ΔUSE1*). Data analysis is presented using the ΔC_T method using a reference gene (see calculations in Material and Methods). As previously mentioned, *rp49* expression levels were not affected between strains therefore this gene was used as endogenous control. Regarding gfp-polo, there is a strong difference between the total *polo* mRNA expression levels and those of the longer isoform (~27-fold) (Figure 10). Since total levels correspond to pA1+pA2 mRNAs

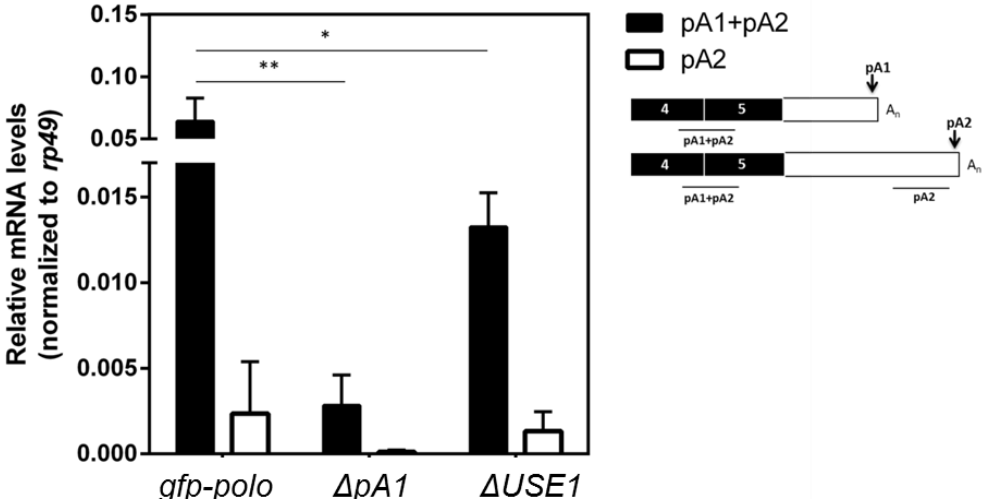


Figure 10. *polo* mRNA isoforms expression levels in *gfp-polo*, *ΔpA1* and *ΔUSE1* transgenic lines. RT–qPCR quantification was performed using real-time PCR oligonucleotides positioned as shown on the right of the graph. Error bars show SEM from at least three independent experiments for *ΔpA1* and *ΔUSE1* and two individual experiments for *gfp-polo*.

and the expression levels of pA2 mRNA are reduced, we can conclude that the pA1 mRNA isoform is the most expressed, which is in agreement with previous observations (58). When compared to gfp-polo, total *polo* levels in *ΔpA1* flies are dramatically reduced (~23-fold), with a strong reduction in *polo* pA1 mRNA as expected, since the pA1 signal is mutated in this transgenic line. Importantly, there was also a reduction in total *polo* levels in the *ΔUSE1* flies when compared to gfp-polo that, although not as pronounced as in *ΔpA1*, still corresponded to a ~5-fold decrease. The levels of the longer isoform did not suffer any significant change (Figure 10), indicating that pA1 usage is compromised in both *ΔpA1* and *ΔUSE1* flies.

2. USE-binding proteins influence *polo* pA site selection

Having established that USE1 deletion correlates with a decrease in *polo* pA1 mRNA isoform production, we then addressed the role of the proteins that bind *polo* USE1: hnRNP C, HuR and PTB. Using RNAi-mediated silencing, we knockdown their correspondent *Drosophila* homologues *elav* and *heph* mRNAs in S2 cells and asked if this affected the production of *polo* mRNA isoforms. Oligonucleotides were designed and optimized to knockdown *cg42458,* the *Drosophila* homologue of hnRNP C, but this experiment is still ongoing.

A. *In vitro* transcriptions and real-time PCR (qPCR) optimization

Oligonucleotides were designed to amplify *DsRed,* (used as a control to distinguish sequence-specific silencing from non-specific effects in the RNAi experiment), *cg42458*, *elav* and *heph*, whose products were subsequently used as DNA template for *in vitro* transcriptions. The first step was to amplify DNA with opposing T7 promoter sequence at the 5'end of each strand (Figure 11A). After DNA amplification, the fragments were purified (Figure 11B) using the QIAquick PCR Purification Kit.

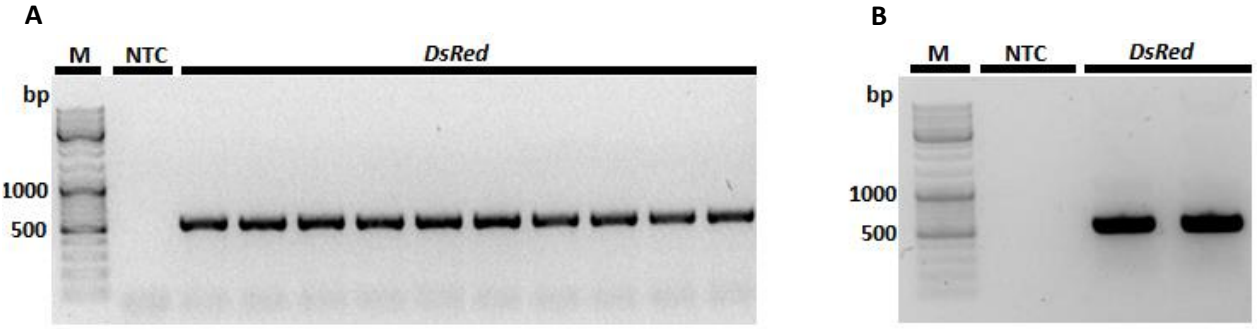


Figure 11. Amplification of *DsRed* using oligonucleotides with opposing T7 promoter sequence at the 5' end of each strand. Total RNA was isolated from *Drosophila* S2 cells and the cDNA was synthetized using random hexamers. (A) *DsRed* fragment before purification (B) *DsRed* fragment after purification M - DNA ladder; NTC- No Template Control.

Finally, DNA fragments were *in vitro* transcribed (Figure 12 lanes 1 and 2), treated with DNase (Figure 12 lanes 3 and 4), cleaned-up and annealed to produce dsRNA. The newly *in vitro* transcribed RNAs (Figure 12 lanes 5 and 6) were subsequently used to knockdown the target genes in *Drosophila* S2 cells.

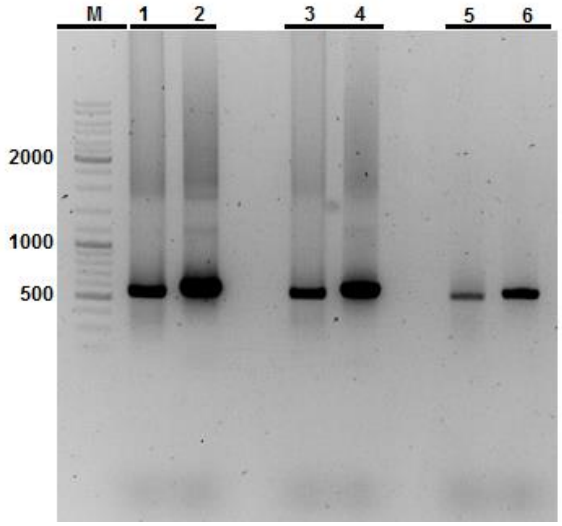


Figure 12. *DsRed* and *elav* dsRNA integrity during and upon *in vitro* transcription 1 and 2 Before the DNase treatment; 3 and 4 After the DNase treatment; 5 and 6 After RNA Clean-up and annealing (See dsRNA synthesis in material and methods).

Simultaneously, oligonucleotides to measure the mRNA expression levels of these target genes by RT-qPCR were designed and optimized (see Material and Methods). All the oligonucleotide pairs had an efficiency of 2, which corresponds to the maximum efficiency.

Secondly, it was necessary to confirm that the *rp49,* used as endogenous control in the experiences with flies, was also the best endogenous control to use in this set of experiments in S2 cells. This was achieved by measuring the C_T values of *rp49* mRNA expression in all of the conditions, control and knockdown samples, respectively.

Table 4. Quantification of *rp49* expression levels in two target samples, *elav* and *heph* knockdown (KD) and comparison with the sample treated with *DsRed* (control).

Target genes	ΔC_T ($C_{T(target)}$-$C_{T(rp49)}$)
***elav* KD**	0.048
***heph* KD**	0.2

As *rp49* expression level proved to be constant across all the samples, *i.e.* ΔC_T is close to zero, it was used as endogenous control in these experiments.

B. RNAi-mediated silencing of *elav* causes a decrease in *polo* pA2 mRNA levels

Function of the USE1-binding proteins in *polo* mRNA isoforms production was addressed using RNAi-mediated silencing of those proteins in *Drosophila* S2 cells. As a control, cells were transformed with a dsRNA of *DsRed* that does not target any known mRNA in the cell (see Material and Methods).

In the first experiment, we seeded S2 cells in12-well plates, with $0.5x10^6$ cells in each well, and when samples were collected after 72 hours, we could not isolate sufficient RNA to perform the analysis. Therefore, the following experiments were all done in 6-well plates. First it was necessary to optimize the concentration of dsRNA and the time of incubation for each knockdown target, as the efficiency of RNAi knockdown varies from gene to gene. The knockdown efficiency in each time-point was determined by RT-qPCR. After 72 hours, 32% of *elav* knockdown was achieved while after 96 hours of incubation, 49±4% of *elav* knockdown was obtained (p value< 0.01) (Fig. 13).

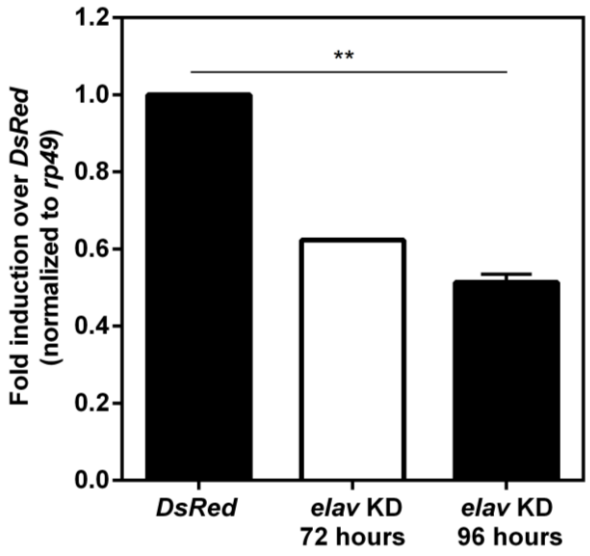


Figure 13. Efficiency of RNAi-mediated silencing of *elav*. *elav* mRNA expression levels were measured after 72 and 96 hours of incubation with dsRNA using RT-qPCR with *elav* specific primers and *rp49* was used as a reference gene. Fold induction over *DsRed* was assessed (DsRed was set to 1). The results include one experiment at 72 hours and three individual experiments at 96 hours (p value< 0. 01).

After an incubation of 96 hours with *elav* dsRNA, both *polo* mRNA isoforms were measured by RT-qPCR and it was observed that the ~50% decrease in *elav* mRNA levels had a strong impact in *polo* mRNA isoforms production.

Figure 14 shows that there was a significant reduction in the levels of the pA2 mRNA isoform, which decreased by 78±30% (p value<0.05) while there was a non-statistically significant increase of total *polo* mRNA levels.

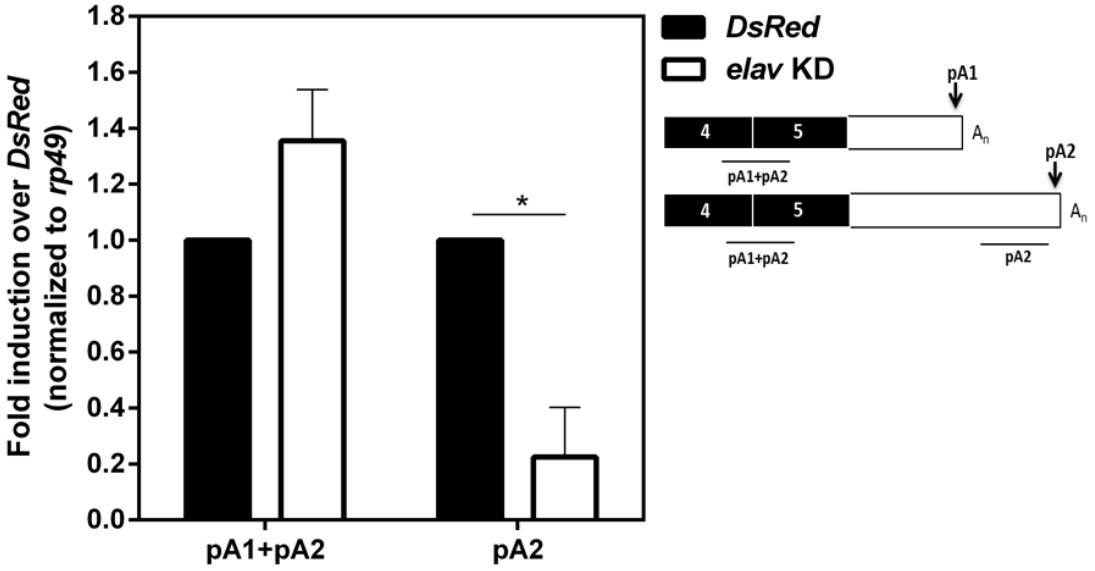


Figure 14. *polo* pA2 mRNA expression is decreased after *elav* knockdown *polo* mRNA isoforms expression levels were measured using RT-qPCR with specific oligonucleotides targeting both *polo* mRNA isoforms and the longer isoform as depicted. *rp49* was used as a reference gene. Fold induction over *DsRed* was displayed (*DsRed* was set to 1). The graph shows the expression levels of *polo* pA1+pA2 mRNA isoforms and the expression levels of *polo* pA2 mRNA, after 96 hours of *elav* knockdown. The results include three individual experiments (p value<0.05).

These results show that *elav* knockdown had an effect in *polo* mRNA isoforms expression levels, specifically by strongly reducing the expression of the longer isoform produced by pA2 usage, indicating that *elav* has a function in *polo* APA.

C. RNAi-mediated silencing of *heph* causes a decrease in *polo* pA2 mRNA levels

Having established the role of Elav in *polo* APA, we then wanted to investigate the role of the other USE-binding proteins. The function of Heph in *polo* mRNA isoforms production was also analysed using RNAi-mediated silencing of *heph* in S2 cells. The concentration of dsRNA and the incubation times used were the same as those for *elav* silencing. Cells were incubated with the correspondent dsRNA, using *DsRed* as a control, collected after 72 and 96 hours and the knockdown efficiency was quantified in both samples. After 72 hours of treatment, 73% of *heph* knockdown was determined while after 96 hours, 89% of *heph* knockdown was achieved (Figure 15).

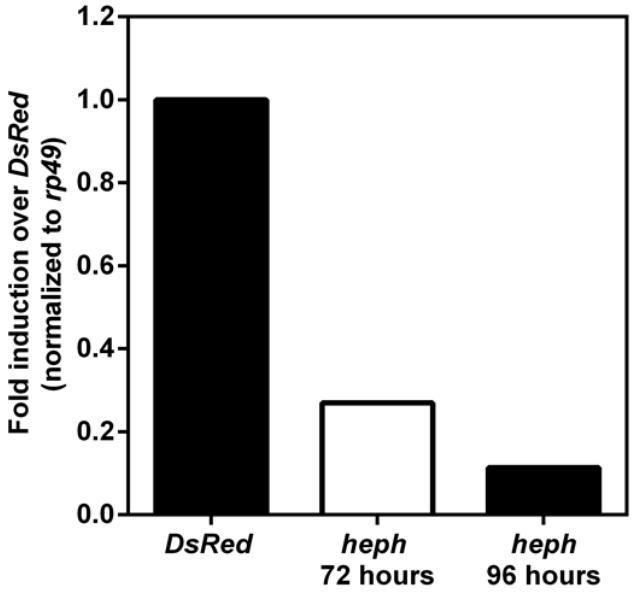


Figure 15. Efficiency of RNAi-mediated silencing of *heph*. *heph* mRNA expression levels were measured after 72 and 96 hours of incubation with dsRNA using RT-qPCR with *heph* specific primers and *rp49* was used as a reference gene. Fold induction over *DsRed* was assessed (DsRed was set to 1). The results include one experiment.

After 96 hours of incubation with *heph* dsRNA, both *polo* mRNA isoforms were quantified by RT-qPCR. It was observed that the 89% decreased in *heph* mRNA levels also had a strong impact in *polo* mRNA production. Figure 16 shows that there was a substantial reduction in the levels of the pA2 mRNA isoform, which decreased by 60% and there was a small increase in total *polo* mRNA levels.

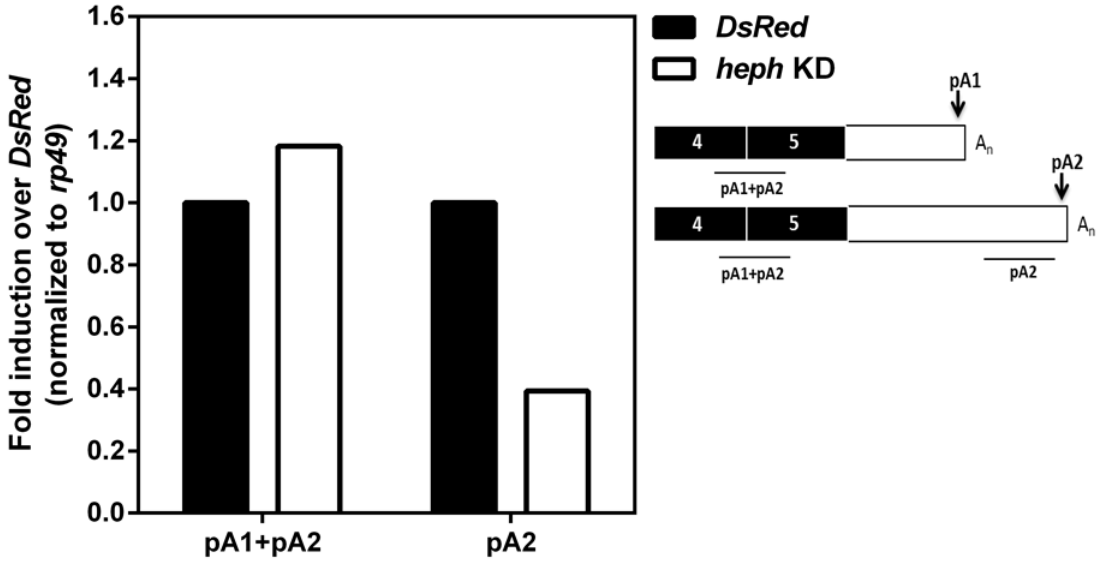


Figure 16. *polo* pA2 mRNA levels are decreased after *heph* knockdown *polo* mRNA isoforms expression levels were measured using RT-qPCR with specific oligonucleotides targeting both polo mRNA isoforms and the longer isoform. *rp49* was used as the reference gene. Fold induction over *DsRed* was displayed (*DsRed* was set to 1). The graph shows the expression

levels of *polo* pA1+pA2 mRNA isoforms and the expression levels of *polo* pA2 mRNA, after 96 hours of *heph* knockdown. The results include one experiment.

.

Our results clearly show that both HuR/Elav and PTB/Heph have a role in *polo* APA in *Drosophila* S2 cells.

IV. Discussion and future perspectives

1. USE1 affects *polo* pA1 mRNA expression levels

Polo expression levels are critical to *Drosophila* development (56) and by APA, *polo* produces two transcripts that have different translational efficiencies. The presence of the pA2 mRNA isoform was found to be essential for fly viability but the pA1 mRNA isoform also showed to be important for the correct formation of the abdomen, indicating that the maintenance of correct transcript levels is crucial to the accurate organism development.

In the first part of this thesis, the impact of USE1 in the expression levels of *polo* mRNA isoforms was assessed by quantifying *polo* mRNA in all the *Drosophila* transgenic lines: w^{1118}; *gfp-polo; polo*^{9}*/TM6B*, w^{1118}; *gfp-polo*Δ*pA1; polo*^{9}*/TM6B* and w^{1118}; *gfp-polo*Δ*USE 1; polo*^{9}*/TM6B*. *ΔpA1* and *ΔUSE1* flies show a similar trend in *polo* mRNA isoforms expression pattern, which is consistent with the similarities between their abdomen phenotype and suggests a role for USE1 in *polo* pA1 site usage. In the flies where pA1 was mutated, a strong reduction in the total amount of *polo* mRNA was observed when compared to gfp-polo flies, but there is no significant change in the pA2 isoform levels detected. This indicates a strong decrease in the pA1 isoform expression levels. This was expected, because the pA1 signal was mutated (ATTAAA to GTTAAC) in order to be non-functional. Neverthless, it is still possible to quantify total *polo* in *ΔpA1* flies. It was already described that important regions for pA signal recognition (USEs) can retain residual levels of polyadenylation activity even in the absence of a functional pA signal sequence (44) and this may contribute to the small total *polo* mRNA levels detected in this strain Also, it was previously described that when a given pA signal is mutated or deleted, an activation of other cryptic pA signals nearby might occur (98). This is also a possible explanation for the total *polo* levels detected in the *ΔpA1* flies since *polo* has a cryptic pA signal between the pA1 and pA2 signals (Moreira, unpublished data). In the future, oligonucleotides should be designed in order to address if this cryptic signal is being used and if it contributes to the total *polo* quantification in these flies.

Although the decrease in total *polo* levels in the *ΔUSE1* flies was not as marked as it was in the *ΔpA1* flies, the trend observed is the same, with a ~5-fold decrease in the total *polo* levels, when compared to *gfp-polo* flies. When expressed, the isoform produced using pA2 appears at very low levels, giving no information about the role of USE1 in its production. Therefore, the strong reduction in total *polo* mRNA levels with the simultaneous unchanged pA2 mRNA isoform levels indicates a strong decrease in the pA1 mRNA isoform expression levels in the absence of USE1.

According to the literature, the main feature of mRNA 3' end formation is the presence of the pA signal, which is still present in *ΔUSE1* flies (11) and can explain the pA1 isoform that is still being produced. It is also important to highlight that several studies indicated that in genes with more than one USE, as in the case of *polo*, these sequences seem to act synergistically (46), therefore it is possible that the other USEs are capable of inducing polyadenylation at the proximal pA signal even if at very low levels. Another interesting aspect to point out is that when measuring the influence of USEs in the 3' end formation, mutation in a given USE is usually responsible for an approximately 50% reduction in polyadenylation (46). Here, we show that deletion of *polo* USE1 is responsible for an 80% decrease in total *polo* mRNA levels.

USEs had already been shown to increase mRNA 3'end formation efficiency in humans, for instance in *cox-2* (46), *IgM secretory* (44) and OASE (99), pro-thrombin and C2 complement (100). Although USEs were only documented in a small number of genes, Legendre and Gautheret showed that these sequences are found in a large number of genes (19). It has been suggested that auxiliary elements play a major role in recruiting or stabilizing the polyadenylation machinery on suboptimal pA signals (44). Based on our results we now propose that the function of USEs is conserved in *Drosophila*.

2. HuR/Elav and PTB/Heph modulate *polo* APA

It had been previously shown that HuR, PTB, hnRNP C and CstF-64 bind to polo USE1, by UV cross-linking assays, using HeLa cells nuclear extracts, followed by immunoprecipitations. (61). The assays were conducted using HeLa cells nuclear extracts due to the lack of a suitable *in vitro Drosophila* system. Therefore, in the second part of this thesis, the function of HuR/Elav and PTB/Heph in *polo* mRNA isoforms expression levels was investigated using RNAi methodologies in *Drosophila* S2 cells.

Our results clearly show that the ~50% *elav* knockdown is responsible for a 78% decrease in *polo* pA2 mRNA isoform expression levels. Since there was no significant difference in the total amount of *polo* mRNA even with a strong decrease in the levels of the longer isoform, this suggests an increase in the level of the shorter isoform. This result indicates that Elav has a negative effect in *polo* pA1 recognition and consequently, influences in pA2 usage. It is possible that by blocking pA1, Elav allows RNA polymerase II (PolII) to read-through resulting in more production of the longer isoform. In fact, the inhibition of mRNA 3' end processing through Elav binding at the vicinity of a pA site was already described in the fly (69). Moreover, it has been recently shown that Elav inhibits proximal pA signals to coordinate the expression of 3' UTR extensions observed during neural development (66). The *heph* knockdown induced a 60% reduction in pA2 mRNA

levels, with an 18% increase in the total amount of *polo* mRNA, which also indicates an increase in the pA1 isoform.

Remarkably, depletion of both proteins has a similar effect in the expression of *polo* mRNA isoforms, which indicate that these proteins might be assembled in a complex. If this is the case, depletion of either protein inhibits the protein complex assembly and binding to the USE1, ultimately resulting in a decrease in the levels of the longer isoforms.

USEs were described as being capable of stabilizing the C/P machinery protein complexes (2, 101), enabling a higher mRNA 3'end efficiency. Whenever Elav and/or Heph are expressed, their interaction with the USE1 would supress the first pA signal and allow read-through of the elongation complexes, resulting in the production of the longer isoform (Figure 17).

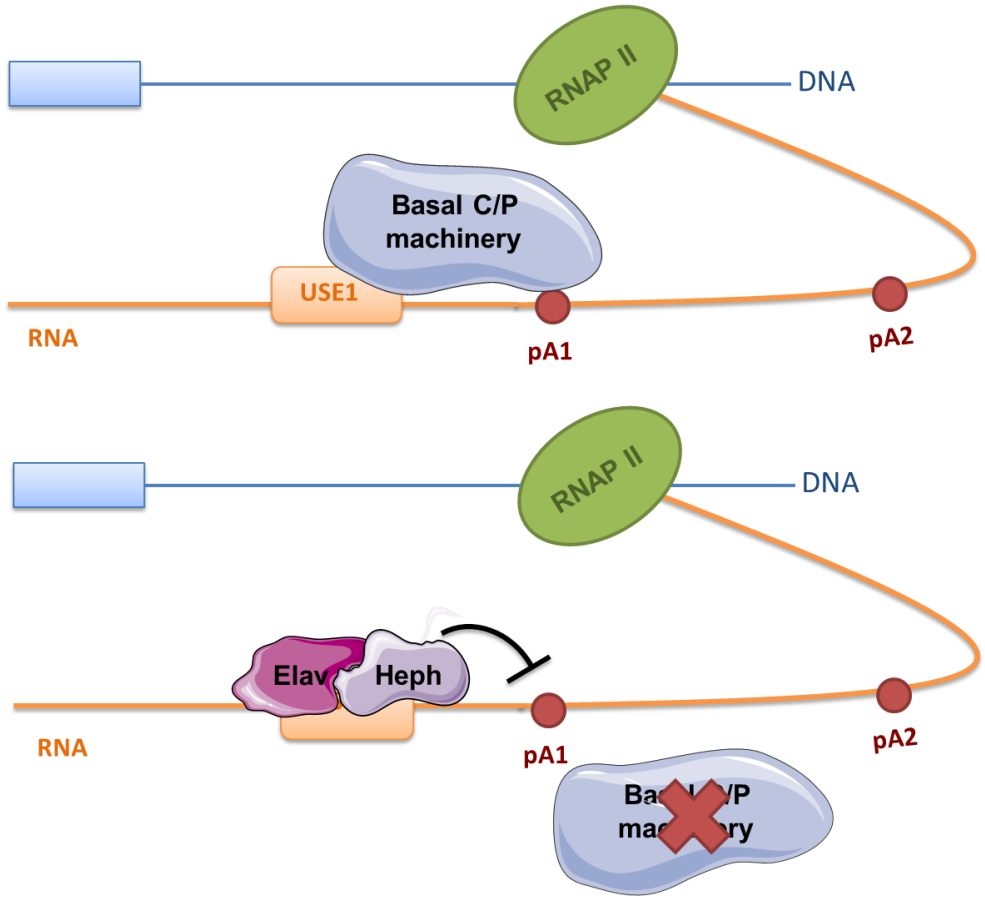


Figure 17. Schematic representation of the proposed effect of Elav and Heph in the usage of the two *polo* pA signals. Elav and Heph bind to U-rich upstream sequence element (USE1) in *polo* mRNA in the vicinity of the proximal pA signal and may play an inhibitory role in pA1 recognition by the basal C/P machinery, in a given cellular environment.

A possible explanation for the results presented in this thesis could be that Elav and/or Heph bind to the USE1 and compete with the basal C/P machinery impairing the assembly of its core protein complexes and as such preventing 3' end formation at pA1. As a consequence, usage of the proximal pA signal would decrease, favouring the production of

the longer mRNA isoform. The levels of this mRNA isoform determined were consistently very low in adult flies, therefore, in the future, it will be necessary to test a system with higher levels of the longer isoform. Since *polo* is highly expressed in the brain during the third instar larval stage, neuroblasts might provide such a system (58, 102), as well as the abdominal histoblasts cells, as pA2 levels are critical for their proliferation (56). Neuroblasts or histoblasts of the transgenic flies could then be used to show that in a system with higher levels of pA2, *ΔUSE1* flies will display a reduction in *polo* pA2 mRNA isoform levels. These results would also support a role for USE1 in pA2 signal usage apart from its role in *polo* 3' end formation at the proximal pA site. Interestingly, Pedro Pinto observed that CstF-64, an important member of the basal C/P machinery, binds upstream of the *polo* pA1.

Competition between USE-binding proteins and C/P machinery was also found to occur in the C2 complement (103), which has an USE that also binds PTB. Castelo-Branco *et al.* showed that PTB competes with CstF for binding to the C2 complement USE sequence. As a consequence, high levels of PTB displace CstF binding and cause a negative effect in polyadenylation (42). Interestingly, PTB/Heph also binds *polo* USE1, suggesting that a similar mechanisms could occur. Additionally, it has been previously shown that interactions between Hu proteins and C/P core machinery prevent the interaction between the CstF-64 subunit and the RNA (70). It will be interesting to further explore this hypothesis, since *Drosophila's* Elav can induce 3'extension in a variety of cell types (2). Since pA2 mRNA has a higher translation efficiency (56), competition for USE1 binding could represent a possible mechanism through which the recognition and usage of the distal pA site is favoured to ensure the required Polo levels for fly survival (56), in a given cellular environment. The results obtained are in agreement with this model, since *elav* or *heph* depletion in *Drosophila* S2 cells by RNAi, reduced the levels of the longer isoform and increased the shorter isoform.

We cannot rule out the possibility of an RNA-binding protein to be differentially regulating the mRNA stability of both transcripts. In fact, hnRNP C, which is involved in mRNA stability (90) also binds USE1. The role of this protein in *polo* polyadenylation pattern could be inferred via the quantification of each *polo* isoforms' stability upon hnRNP C depletion via RNAi in *Drosophila* Schneider cells.

The results obtained throughout this thesis show that *polo* USE1 modulates APA *in vivo*. In *Drosophila*, the proteins that bind *polo* USE1, Elav and Heph, were shown in this work to be important to maintain the correct levels of each transcript. These, in turn are known to play an important role in the production of the appropriate Polo protein levels, ultimately promoting the normal development of the fly (1). It is likely that *polo* APA is under the influence of these USE-binding proteins expression pattern. This is one of the first

reports that provide evidence for the role of USEs *in vivo* revealing a new level of conservation in APA regulation between mammals and *Drosophila.*

Funding

This work was supported by FEDER through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT– Fundação para a Ciência e a Tecnologia under the project FCOMP-01-0124-FEDER-028252 (PTDC/BEX-BCM/0468/2012).

V. References

1. Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA processing with transcription. Cell. 2002;108(4):501-12.
2. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiology and molecular biology reviews : MMBR. 1999;63(2):405-45.
3. Bentley DL. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Current opinion in cell biology. 2005;17(3):251-6.
4. Proudfoot NJ, Brownlee GG. 3' non-coding region sequences in eukaryotic messenger RNA. Nature. 1976;263(5574):211-4.
5. Conway L, Wickens M. Analysis of mRNA 3' end formation by modification interference: the only modifications which prevent processing lie in AAUAAA and the poly(A) site. The EMBO journal. 1987;6(13):4177-84.
6. Sheets MD, Ogg SC, Wickens MP. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic acids research. 1990;18(19):5799-805.
7. Wilusz J, Pettine SM, Shenk T. Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. Nucleic acids research. 1989;17(10):3899-908.
8. Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. Genome research. 2000;10(7):1001-10.
9. MacDonald CC, Redondo JL. Reexamining the polyadenylation signal: were we wrong about AAUAAA? Molecular and cellular endocrinology. 2002;190(1-2):1-8.
10. Retelska D, Iseli C, Bucher P, Jongeneel CV, Naef F. Similarities and differences of polyadenylation signals in human and fly. BMC genomics. 2006;7:176.
11. Wickens M, Stephenson P. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. Science (New York, NY). 1984;226(4678):1045-51.
12. Qian ZW, Wilusz J. An RNA-binding protein specifically interacts with a functionally important domain of the downstream element of the simian virus 40 late polyadenylation signal. Molecular and cellular biology. 1991;11(10):5312-20.
13. Ryner LC, Takagaki Y, Manley JL. Sequences downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation in vitro both directly and indirectly. Molecular and cellular biology. 1989;9(4):1759-71.
14. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. Nucleic acids research. 2003;31(5):1375-86.
15. Chen F, Wilusz J. Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. Nucleic acids research. 1998;26(12):2891-8.
16. Chen F, MacDonald CC, Wilusz J. Cleavage site determinants in the mammalian polyadenylation signal. Nucleic acids research. 1995;23(14):2614-20.

17. Gehring NH, Frede U, Neu-Yilik G, Hundsdoerfer P, Vetter B, Hentze MW, et al. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. Nature genetics. 2001;28(4):389-92.
18. Danckwardt S, Hartmann K, Katz B, Hentze MW, Levy Y, Eichele R, et al. The prothrombin 20209 C-->T mutation in Jewish-Moroccan Caucasians: molecular analysis of gain-of-function of 3' end processing. Journal of thrombosis and haemostasis : JTH. 2006;4(5):1078-85.
19. Legendre M, Gautheret D. Sequence determinants in human polyadenylation site selection. BMC genomics. 2003;4(1):7.
20. Miura P, Sanfilippo P, Shenker S, Lai EC. Alternative polyadenylation in the nervous system: to what lengths will 3' UTR extensions take us? BioEssays : news and reviews in molecular, cellular and developmental biology. 2014;36(8):766-77.
21. Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. The EMBO journal. 2004;23(3):616-26.
22. Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes & development. 1995;9(21):2672-83.
23. MacDonald CC, Wilusz J, Shenk T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Molecular and cellular biology. 1994;14(10):6647-54.
24. Brown KM, Gilmartin GM. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Molecular cell. 2003;12(6):1467-76.
25. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes & development. 2005;19(11):1315-27.
26. Christofori G, Keller W. Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. Molecular and cellular biology. 1989;9(1):193-203.
27. Takagaki Y, Ryner LC, Manley JL. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell. 1988;52(5):731-42.
28. Terns MP, Jacob ST. Role of poly(A) polymerase in the cleavage and polyadenylation of mRNA precursor. Molecular and cellular biology. 1989;9(4):1435-44.
29. Takagaki Y, Manley JL. Complex protein interactions within the human polyadenylation machinery identify a novel component. Molecular and cellular biology. 2000;20(5):1515-25.
30. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature. 1997;385(6614):357-61.
31. Hirose Y, Manley JL. RNA polymerase II is an essential mRNA polyadenylation factor. Nature. 1998;395(6697):93-6.
32. Carswell S, Alwine JC. Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. Molecular and cellular biology. 1989;9(10):4248-58.
33. Russnak R, Ganem D. Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. Genes & development. 1990;4(5):764-76.
34. Sanfacon H, Brodmann P, Hohn T. A dissection of the cauliflower mosaic virus polyadenylation signal. Genes & development. 1991;5(1):141-9.
35. DeZazzo JD, Imperiale MJ. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. Molecular and cellular biology. 1989;9(11):4951-61.
36. Cherrington J, Ganem D. Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. The EMBO journal. 1992;11(4):1513-24.
37. Valsamakis A, Zeichner S, Carswell S, Alwine JC. The human immunodeficiency virus type 1 polyadenylylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for

efficient polyadenylylation. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(6):2108-12.
38. DeZazzo JD, Kilpatrick JE, Imperiale MJ. Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation. Molecular and cellular biology. 1991;11(3):1624-30.
39. Brown PH, Tiley LS, Cullen BR. Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences. Journal of virology. 1991;65(6):3340-3.
40. Graveley BR, Gilmartin GM. A common mechanism for the enhancement of mRNA 3' processing by U3 sequences in two distantly related lentiviruses. Journal of virology. 1996;70(3):1612-7.
41. Moreira A, Wollerton M, Monks J, Proudfoot NJ. Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. The EMBO journal. 1995;14(15):3809-19.
42. Castelo-Branco P, Furger A, Wollerton M, Smith C, Moreira A, Proudfoot N. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. Molecular and cellular biology. 2004;24(10):4174-83.
43. Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ. The upstream sequence element of the C2 complement poly (A) signal activates mRNA 3' end formation by two distinct mechanisms. Genes & development. 1998;12(16):2522-34.
44. Phillips C, Virtanen A. The murine IgM secretory poly(A) site contains dual upstream and downstream elements which affect polyadenylation. Nucleic acids research. 1997;25(12):2344-51.
45. Natalizio BJ, Muniz LC, Arhin GK, Wilusz J, Lutz CS. Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. The Journal of biological chemistry. 2002;277(45):42733-40.
46. Hall-Pogar T, Zhang H, Tian B, Lutz CS. Alternative polyadenylation of cyclooxygenase-2. Nucleic acids research. 2005;33(8):2565-79.
47. Danckwardt S, Gehring NH, Neu-Yilik G, Hundsdoerfer P, Pforsich M, Frede U, et al. The prothrombin 3'end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. Blood. 2004;104(2):428-35.
48. Danckwardt S, Kaufmann I, Gentzel M, Foerstner KU, Gantzert AS, Gehring NH, et al. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. The EMBO journal. 2007;26(11):2658-69.
49. Lutz CS. Alternative polyadenylation: a twist on mRNA 3' end formation. ACS chemical biology. 2008;3(10):609-17.
50. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nature reviews Genetics. 2013;14(7):496-506.
51. Liao B, Hu Y, Brewer G. Competitive binding of AUF1 and TIAR to MYC mRNA controls its translation. Nature structural & molecular biology. 2007;14(6):511-8.
52. Tiedje C, Ronkina N, Tehrani M, Dhamija S, Laass K, Holtmann H, et al. The p38/MK2-driven exchange between tristetraprolin and HuR regulates AU-rich element-dependent translation. PLoS genetics. 2012;8(9):e1002977.
53. Sunkel CE, Glover DM. polo, a mitotic mutant of Drosophila displaying abnormal spindle poles. Journal of cell science. 1988;89 (Pt 1):25-38.
54. Lowery DM, Lim D, Yaffe MB. Structure and function of Polo-like kinases. Oncogene. 2005;24(2):248-59.
55. Zitouni S, Nabais C, Jana SC, Guerrero A, Bettencourt-Dias M. Polo-like kinases: structural variations lead to multiple functions. Nature reviews Molecular cell biology. 2014;15(7):433-52.
56. Pinto PA, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, et al. RNA polymerase II kinetics in polo polyadenylation signal selection. The EMBO journal. 2011;30(12):2431-44.

57. Weaver RF. Molecular Biology. New York: McGraw-Hill; 2012.
58. Llamazares S, Moreira A, Tavares A, Girdham C, Spruce BA, Gonzalez C, et al. polo encodes a protein kinase homolog required for mitosis in Drosophila. Genes & development. 1991;5(12a):2153-65.
59. Moutinho-Santos T, Sampaio P, Amorim I, Costa M, Sunkel CE. In vivo localisation of the mitotic POLO kinase shows a highly dynamic association with the mitotic apparatus during early embryogenesis in Drosophila. Biology of the cell / under the auspices of the European Cell Biology Organization. 1999;91(8):585-96.
60. Donaldson MM, Tavares AA, Ohkura H, Deak P, Glover DM. Metaphase arrest with centromere separation in polo mutants of Drosophila. The Journal of cell biology. 2001;153(4):663-76.
61. Pinto PAB. The role of *Drosophila melanogster polo* 3' untranslated region in gene expression [Dissertação de Doutoramento em Ciências Biomédicas]. Porto: Universidade do Porto; 2008.
62. Robinow S, Campos AR, Yao KM, White K. The elav gene product of Drosophila, required in neurons, has three RNP consensus motifs. Science (New York, NY). 1988;242(4885):1570-2.
63. Koushika SP, Lisbin MJ, White K. ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform. Current biology : CB. 1996;6(12):1634-41.
64. Koushika SP, Soller M, White K. The neuron-enriched splicing pattern of Drosophila erect wing is dependent on the presence of ELAV protein. Molecular and cellular biology. 2000;20(5):1836-45.
65. Samson ML. Evidence for 3' untranslated region-dependent autoregulation of the Drosophila gene encoding the neuronal nuclear RNA-binding protein ELAV. Genetics. 1998;150(2):723-33.
66. Hilgers V, Lemke SB, Levine M. ELAV mediates 3' UTR extension in the Drosophila nervous system. Genes & development. 2012;26(20):2259-64.
67. Rogulja-Ortmann A, Picao-Osorio J, Villava C, Patraquim P, Lafuente E, Aspden J, et al. The RNA-binding protein ELAV regulates Hox RNA processing, expression and function within the Drosophila nervous system. Development. 2014;141(10):2046-56.
68. Lisbin MJ, Qiu J, White K. The neuron-specific RNA-binding protein ELAV regulates neuroglian alternative splicing in neurons and binds directly to its pre-mRNA. Genes & development. 2001;15(19):2546-61.
69. Soller M, White K. ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA. Genes & development. 2003;17(20):2526-38.
70. Zhu H, Zhou HL, Hasman RA, Lou H. Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. The Journal of biological chemistry. 2007;282(4):2203-10.
71. Kamath RV, Leary DJ, Huang S. Nucleocytoplasmic shuttling of polypyrimidine tract-binding protein is uncoupled from RNA export. Molecular biology of the cell. 2001;12(12):3808-20.
72. Carstens RP, Wagner EJ, Garcia-Blanco MA. An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein. Molecular and cellular biology. 2000;20(19):7388-400.
73. Chan RC, Black DL. The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. Molecular and cellular biology. 1997;17(8):4667-76.
74. Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Molecular cell. 2000;5(6):949-57.
75. Cote J, Dupuis S, Wu JY. Polypyrimidine track-binding protein binding downstream of caspase-2 alternative exon 9 represses its inclusion. The Journal of biological chemistry. 2001;276(11):8535-43.

76. Lou H, Helfman DM, Gagel RF, Berget SM. Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. Molecular and cellular biology. 1999;19(1):78-85.
77. Mulligan GJ, Guo W, Wormsley S, Helfman DM. Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. The Journal of biological chemistry. 1992;267(35):25480-7.
78. Perez I, McAfee JG, Patton JG. Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. Biochemistry. 1997;36(39):11881-90.
79. Southby J, Gooding C, Smith CW. Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of alpha-actinin mutally exclusive exons. Molecular and cellular biology. 1999;19(4):2699-711.
80. Zhang L, Liu W, Grabowski PJ. Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. Rna. 1999;5(1):117-30.
81. Kim YK, Hahm B, Jang SK. Polypyrimidine tract-binding protein inhibits translation of bip mRNA. Journal of molecular biology. 2000;304(2):119-33.
82. Valcarcel J, Gebauer F. Post-transcriptional regulation: the dawn of PTB. Current biology : CB. 1997;7(11):R705-8.
83. Millevoi S, Decorsiere A, Loulergue C, Iacovoni J, Bernat S, Antoniou M, et al. A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. Nucleic acids research. 2009;37(14):4672-83.
84. Dansereau DA, Lunke MD, Finkielsztein A, Russell MA, Brook WJ. Hephaestus encodes a polypyrimidine tract binding protein that regulates Notch signalling during wing development in Drosophila melanogaster. Development. 2002;129(24):5553-66.
85. Besse F, Lopez de Quinto S, Marchand V, Trucco A, Ephrussi A. Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. Genes & development. 2009;23(2):195-207.
86. Burd CG, Swanson MS, Gorlach M, Dreyfuss G. Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(24):9788-92.
87. Tran H, Maurer F, Nagamine Y. Stabilization of urokinase and urokinase receptor mRNAs by HuR is linked to its cytoplasmic accumulation induced by activated mitogen-activated protein kinase-activated protein kinase 2. Molecular and cellular biology. 2003;23(20):7177-88.
88. Esnault S, Malter JS. Hyaluronic acid or TNF-alpha plus fibronectin triggers granulocyte macrophage-colony-stimulating factor mRNA stabilization in eosinophils yet engages differential intracellular pathways and mRNA binding proteins. Journal of immunology (Baltimore, Md : 1950). 2003;171(12):6780-7.
89. Rajagopalan LE, Westmark CJ, Jarzembowski JA, Malter JS. hnRNP C increases amyloid precursor protein (APP) production by stabilizing APP mRNA. Nucleic acids research. 1998;26(14):3418-23.
90. Shetty S. Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. Molecular and cellular biochemistry. 2005;272(1-2):107-18.
91. Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. Annual review of biochemistry. 1993;62:289-321.
92. Wilusz J, Shenk T. A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. Molecular and cellular biology. 1990;10(12):6397-407.
93. Christian KJ, Lang MA, Raffalli-Mathieu F. Interaction of heterogeneous nuclear ribonucleoprotein C1/C2 with a novel cis-regulatory element within p53 mRNA as a response to cytostatic drug treatment. Molecular pharmacology. 2008;73(5):1558-67.
94. Sambrook J, Russell D. Molecular Cloning: A Laboratory Manual2001.

95. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3 - new capabilities and interfaces. Nucleic acids research. 2012;40(15):e115.
96. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics (Oxford, England). 2007;23(10):1289-91.
97. Kulkarni MM, Booker M, Silver SJ, Friedman A, Hong P, Perrimon N, et al. Evidence of off-target effects associated with long dsRNAs in Drosophila melanogaster cell-based assays. Nature methods. 2006;3(10):833-8.
98. Costessi L, Porro F, Iaconcig A, Nedeljkovic M, Muro AF. Characterization of the distal polyadenylation site of the ss-adducin (Add2) pre-mRNA. PloS one. 2013;8(3):e58879.
99. Aissouni Y, Perez C, Calmels B, Benech PD. The cleavage/polyadenylation activity triggered by a U-rich motif sequence is differently required depending on the poly(A) site location at either the first or last 3'-terminal exon of the 2'-5' oligo(A) synthetase gene. The Journal of biological chemistry. 2002;277(39):35808-14.
100. Moreira A, Wollerton M, Monks J, Proudfoot NJ. Upstream sequence elements enhance poly (A) site efficiency of the C2 complement gene and are phylogenetically conserved. The EMBO journal. 1995;14(15):3809.
101. Lutz CS, Murthy KG, Schek N, O'Connor JP, Manley JL, Alwine JC. Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes & development. 1996;10(3):325-37.
102. Tavares AA, Glover DM, Sunkel CE. The conserved mitotic kinase polo is regulated by phosphorylation and has preferred microtubule-associated substrates in Drosophila embryo extracts. The EMBO journal. 1996;15(18):4873-83.
103. Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. Genes & development. 1998;12(16):2522-34.